

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2004, 11:13:25 ; Search time 22.2651 Seconds
(without alignments)
83.400 Million cell updates/sec

Title: US-10-783-095-11
Perfect score: 28
Sequence: 1 LAKEMQALCAVQAPNTCATQGGNIK 28

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 478139 seqs, 66318000 residues

Word size : 0

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	28	100.0	294	2	US-08-468-576B-18
2	28	100.0	294	2	US-08-468-579B-18
3	28	100.0	294	3	US-08-468-577B-18
4	28	100.0	548	2	US-08-468-576B-19
5	28	100.0	548	2	US-08-468-579B-19
6	28	100.0	548	3	US-08-468-577B-19
7	28	100.0	979	3	US-08-514-213A-2
8	28	100.0	979	4	US-09-015-399-5
9	24	85.7	24	4	US-09-552-802B-22
10	19	67.9	19	4	US-09-552-802B-20
11	18	64.3	18	4	US-09-552-802B-21
12	17	60.7	17	4	US-09-552-802B-19
13	16	57.1	16	4	US-09-552-802B-18
14	13	46.4	13	4	US-09-552-802B-17
15	10	35.7	298	3	US-08-811-481-4
16	10	35.7	298	4	US-09-876-527-4
17	10	35.7	276	3	US-08-811-481-7
18	10	35.7	376	4	US-09-876-527-7
19	10	35.7	641	3	US-08-811-481-2
20	10	35.7	641	4	US-09-876-527-2
21	10	35.7	723	2	US-08-548-159-5
22	10	35.7	818	3	US-08-811-481-22
23	10	35.7	818	4	US-09-876-527-22
24	10	35.7	969	2	US-08-548-159-1
25	10	35.7	986	2	US-08-548-159-3
26	10	35.7	1001	3	US-08-884-569A-2
27	10	35.7	1012	3	US-08-811-481-16

Sequence 16, Appl
Sequence 49, Appl
Sequence 9193, Ap
Sequence 9959, Ap
Sequence 10543, A
Sequence 11154, A
Sequence 11319, A
Sequence 11469, A
Sequence 11572, A
Sequence 11873, A
Sequence 12042, A
Sequence 12617, A
Patent No. 5187077
Sequence 43882, A
Sequence 61551, A
Sequence 24690, A
Sequence 15, Appl
Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-468-576B-18
Sequence 18, Application US/08468576B
Patent No. 5955345
GENERAL INFORMATION:
APPLICANT: Rabin, Daniel
TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Spring Kramer Schaefer & Briscoe
STREET: 660 White Plains Road
CITY: Tarrytown
STATE: New York
COUNTRY: USA
ZIP: 10591-5144
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: System 7.5
SOFTWARE: WordPerfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/468,576B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/239,276
FILING DATE: 05-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/872,646
FILING DATE: 08-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/715,181
FILING DATE: 14-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/441,703
FILING DATE: 04-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/312,543
FILING DATE: 17-FEB-1989
ATTORNEY/AGENT INFORMATION:
NAME: Kurt G. Briscoe
REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: MDI 251.7-KGB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 332-1700
TELEFAX: (914) 332-1844
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 294 amino acids

TYPE: amino acid
TOPOLOGY: linear
US-08-468-576B-18

Query Match 100.0%; Score 28; DB 2; Length 294;
Best Local Similarity 100.0%; Pred. No. 6.3e-22;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LAKEWQALCAVQAEPTCATAGGEGNIK 28
DB 67 LAKEWQALCAVQAEPTCATAGGEGNIK 94

RESULT 2

US-08-468-579B-18
; Sequence 18, Application US/08468579B
; Patent No. 5981700
; GENERAL INFORMATION:
; APPLICANT: Rabin, Daniel
; TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sprung Kramer Schaefer & Briscoe
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: System 7.5
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,579B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/239,276
; FILING DATE: 05-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,646
; FILING DATE: 08-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/715,181
; FILING DATE: 14-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/441,703
; FILING DATE: 04-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/312,543
; FILING DATE: 04-DEC-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Kurt G. Briscoe
; REGISTRATION NUMBER: 33,141
; REFERENCE/DOCKET NUMBER: MDI 251.5-KGB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914) 332-1700
; TELEFAX: (914) 332-1844
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 294 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-468-579B-18

Query Match 100.0%; Score 28; DB 2; Length 294;
Best Local Similarity 100.0%; Pred. No. 6.3e-22;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LAKEWQALCAVQAEPTCATAGGEGNIK 28
DB 67 LAKEWQALCAVQAEPTCATAGGEGNIK 94

RESULT 3

US-08-468-577B-18
; Sequence 18, Application US/08468577B
; Patent No. 6001804
; GENERAL INFORMATION:
; APPLICANT: Rabin, Daniel
; TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sprung Kramer Schaefer & Briscoe
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: System 7.5
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,577B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/239,276
; FILING DATE: 05-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,646
; FILING DATE: 08-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/715,181
; FILING DATE: 14-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/441,703
; FILING DATE: 04-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/312,543
; FILING DATE: 17-FEB-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Kurt G. Briscoe
; REGISTRATION NUMBER: 33,141
; REFERENCE/DOCKET NUMBER: MDI 251.8-KGB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914) 332-1700
; TELEFAX: (914) 332-1844
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 294 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-468-577B-18

Query Match 100.0%; Score 28; DB 3; Length 294;
Best Local Similarity 100.0%; Pred. No. 6.3e-22;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LAKEWQALCAVQAEPTCATAGGEGNIK 28
DB 67 LAKEWQALCAVQAEPTCATAGGEGNIK 94

RESULT 4

US-08-468-576B-19
; Sequence 19, Application US/08468576B
; Patent No. 5955345
; GENERAL INFORMATION:
; APPLICANT: Rabin, Daniel
; TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2004, 11:13:25 ; Search time 8.74699 Seconds
(without alignments)
83.400 Million cell updates/sec

Title: US-10-783-095-9
Perfect score: 11
Sequence: 1 QPLALEGSLQK 11

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 478139 seqs, 66318000 residues

Word size : 0

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/FACTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	11	100.0	23	3	US-08-105-904B-10
2	11	100.0	23	3	US-08-114-877A-10
3	11	100.0	23	3	US-08-537-928-10
4	11	100.0	23	4	US-09-633-269-10
5	11	100.0	27	2	US-08-508-664-17
6	11	100.0	29	4	US-09-676-787-2
7	11	100.0	32	6	5514646-5
8	11	100.0	33	4	US-09-639-667-24
9	11	100.0	33	6	5514646-4
10	11	100.0	35	1	US-08-160-376A-3
11	11	100.0	35	1	US-08-389-487-6
12	11	100.0	35	2	US-09-134-836-3
13	11	100.0	35	3	US-09-386-303A-3
14	11	100.0	35	4	US-09-815-229-4
15	11	100.0	35	4	US-09-947-563-3
16	11	100.0	35	6	5457066-1
17	11	100.0	54	4	US-09-815-229-13
18	11	100.0	86	1	US-07-715-183C-1
19	11	100.0	86	4	US-09-477-924-2
20	11	100.0	86	4	US-09-723-981-2
21	11	100.0	86	4	US-09-723-896-2
22	11	100.0	86	4	US-09-878-380-1
23	11	100.0	90	1	US-08-030-731A-43
24	11	100.0	91	4	US-09-676-787-7
25	11	100.0	96	2	US-09-134-836-4
26	11	100.0	96	2	US-09-134-836-5
27	11	100.0	96	3	US-09-386-303A-4

28 11 100.0 96 3 US-09-386-303A-5 Sequence 5, Appli
29 11 100.0 96 4 US-09-947-563-4 Sequence 4, Appli
30 11 100.0 96 4 US-09-947-563-5 Sequence 5, Appli
31 11 100.0 97 1 US-08-160-376A-4 Sequence 4, Appli
32 11 100.0 97 1 US-08-389-487-7 Sequence 7, Appli
33 11 100.0 97 3 US-09-099-307-6 Sequence 6, Appli
34 11 100.0 97 3 US-09-099-307-7 Sequence 7, Appli
35 11 100.0 97 3 US-09-099-307-8 Sequence 8, Appli
36 11 100.0 97 3 US-09-099-307-11 Sequence 11, Appli
37 11 100.0 98 4 US-09-701-968-7 Sequence 7, Appli
38 11 100.0 99 4 US-09-701-968-8 Sequence 8, Appli
39 11 100.0 100 4 US-09-701-968-9 Sequence 9, Appli
40 11 100.0 110 3 US-08-950-720A-11 Sequence 11, Appli
41 11 100.0 110 3 US-08-589-028-2 Sequence 2, Appli
42 11 100.0 110 3 US-08-589-028-4 Sequence 4, Appli
43 11 100.0 110 3 US-08-784-582-2 Sequence 2, Appli
44 11 100.0 110 3 US-08-784-582-4 Sequence 4, Appli
45 11 100.0 110 3 US-08-785-271-2 Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-105-904B-10
; Sequence 10, Application US/08105904B
; Patent No. 6001364
; GENERAL INFORMATION:
; APPLICANT: Rose, Keith
; APPLICANT: Offord, Robin
; TITLE OF INVENTION: HETERO-POLYOXIME COMPOUNDS AND THEIR
; PREPARATION BY PARALLEL ASSEMBLY
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooley Godward Castro Huddleson & Tatum
; STREET: 5 Palo Alto Square, 3000 El Camino Real
; CITY: Palo Alto
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/105,904B
; FILING DATE: 31-AUG-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/057,594
; FILING DATE: 05-MAY-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Neeley, Richard L.
; REGISTRATION NUMBER: 30,092
; REFERENCE/DOCKET NUMBER: ABIC-001/02US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)843-5000
; TELEFAX: (415)857-0663
; TELEX: 380816 Coolypa
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 23 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEICAL: NO
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 1
; OTHER INFORMATION: AOA-Glu
US-08-105-904B-10

Query Match 100.0%; Score 11; DB 3; Length 23;
 Best Local Similarity 100.0%; Pred. No. 3.9e-05;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPLALEGSLOK 11
 |||||
 Db 12 QPLALEGSLOK 22

RESULT 2

US-08-114-877A-10
 ; Sequence 10, Application US/08114877A
 ; Patent No. 6174530
 ; GENERAL INFORMATION:
 ; APPLICANT: Rose, Keith
 ; TITLE OF INVENTION: HOMOGENOUS POLYOXIME COMPOSITIONS AND THEIR
 ; PREPARATION BY PARALLEL ASSEMBLY
 ; NUMBER OF SEQUENCES: 15
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Cooley Godward Castro Huddleson & Tatum
 ; STREET: 5 Palo Alto Square
 ; CITY: Palo Alto
 ; STATE: California
 ; COUNTRY: U.S.A.
 ; ZIP: 94036

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/114,877A
 FILING DATE: 31-AUG-1993

CLASSIFICATION: 424
 PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/057,594
 FILING DATE: 05-MAY-1993

CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:

NAME: Neeley, Richard L.
 REGISTRATION NUMBER: 30,092

REFERENCE/DOCKET NUMBER: ABIC-001/01US
 TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 843 5070
 TELEFAX: (415) 857-0663

TELEX: 380816 COOLEYPA
 INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:
 LENGTH: 23 amino acids

TYPE: amino acid
 TOPOLOGY: linear

MOLECULE TYPE: peptide
 HYPOTHETICAL: NO

NAME/KEY: Modified-site
 LOCATION: 1

OTHER INFORMATION: ROA-Glu
 US-08-114-877A-10

Query Match 100.0%; Score 11; DB 3; Length 23;
 Best Local Similarity 100.0%; Pred. No. 3.9e-05;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPLALEGSLOK 11
 |||||
 Db 12 QPLALEGSLOK 22

RESULT 3

US-08-537-928-10
 ; Sequence 10, Application US/08537928

Patent No. 6217873
 ; GENERAL INFORMATION:
 ; APPLICANT: Keith, Rose
 ; TITLE OF INVENTION: POLYOXIME COMPOSITIONS AND THEIR PREPARATION
 ; FILE REFERENCE: GRFN-001/03US
 ; CURRENT APPLICATION NUMBER: US/08/537,928
 ; EARLIER FILING DATE: 1996-01-05
 ; EARLIER APPLICATION NUMBER: PCT/IB94/00093
 ; EARLIER FILING DATE: 1994-05-05
 ; EARLIER APPLICATION NUMBER: 08/105,904
 ; EARLIER FILING DATE: 1993-08-31
 ; EARLIER APPLICATION NUMBER: 08/114,877
 ; EARLIER FILING DATE: 1993-08-31
 ; NUMBER OF SEQ ID NOS: 30
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 10
 ; LENGTH: 23
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 ; NAME/KEY: SITE
 ; LOCATION: (1)
 ; OTHER INFORMATION: Aminoxyacetyl-Glu
 US-08-537-928-10

Query Match 100.0%; Score 11; DB 3; Length 23;
 Best Local Similarity 100.0%; Pred. No. 3.9e-05;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPLALEGSLOK 11
 |||||
 Db 12 QPLALEGSLOK 22

RESULT 4

US-09-633-269-10
 ; Sequence 10, Application US/09633269

Patent No. 6663869
 ; GENERAL INFORMATION:

APPLICANT: Keith, Rose
 ; TITLE OF INVENTION: POLYOXIME COMPOSITIONS AND THEIR PREPARATION

FILE REFERENCE: GRFN-001/03US
 ; CURRENT APPLICATION NUMBER: US/09/633,269

CURRENT FILING DATE: 2000-08-04
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/08/537,928

PRIOR FILING DATE: EARLIER FILING DATE: 1996-01-05
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PCT/IB94/00093

PRIOR FILING DATE: EARLIER FILING DATE: 1994-05-05
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 08/105,904

PRIOR FILING DATE: EARLIER FILING DATE: 1993-08-31
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 08/114,877

PRIOR FILING DATE: EARLIER FILING DATE: 1993-08-31
 ; NUMBER OF SEQ ID NOS: 30

SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 10

LENGTH: 23
 ; TYPE: PRT

ORGANISM: Artificial Sequence
 ; FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 ; NAME/KEY: SITE
 ; LOCATION: (1)
 ; OTHER INFORMATION: Aminoxyacetyl-Glu
 US-09-633-269-10

Query Match 100.0%; Score 11; DB 4; Length 23;
 Best Local Similarity 100.0%; Pred. No. 3.9e-05;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2004, 06:03:19 ; Search time 38.9313 Seconds
(without alignments)
101.620 Million cell updates/sec

Title: US-10-783-095-9
Perfect score: 53
Sequence: 1 QPLAEGSLQX 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_23Sep04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53	100.0	12	2 AAW06808	Aaw06808 Single ch
2	53	100.0	18	8 ADI29027	Adi29027 Proinsuli
3	53	100.0	23	2 AAR60842	Aar60842 AoA-PepC
4	53	100.0	27	2 AAW81858	Aaw81858 Human pro
5	53	100.0	29	4 AAB70847	Aab70847 Human pro
6	53	100.0	29	8 ADN34533	Adn34533 Preproins
7	53	100.0	32	1 AAF50011	Aaf50011 Sequence
8	53	100.0	33	2 AAR83840	Aar83840 Part of i
9	53	100.0	35	1 AAP40830	Aap40830 Sequence
10	53	100.0	35	2 AAR68897	Aar68897 Human ins
11	53	100.0	35	2 AAY08003	Aay08003 Human ins
12	53	100.0	35	2 AAO17829	Aao17829 Human ins
13	53	100.0	35	4 AAG65878	Aag65878 Human pro
14	53	100.0	54	4 AAG65887	Aag65887 Fragment
15	53	100.0	63	2 AAW06807	Aaw06807 Single ch
16	53	100.0	64	2 AAR38432	Aar38432 Split (64)
17	53	100.0	64	2 AAR38433	Aar38433 Split (64)
18	53	100.0	65	2 AAR07660	Aar07660 Human ins
19	53	100.0	65	2 AAW34587	Aaw34587 Insulin c
20	53	100.0	66	2 AAW06809	Aaw06809 Met-Arg-M
21	53	100.0	72	4 AAUC0265	Aau0265 Synthetic
22	53	100.0	83	1 AAP70639	Aap70639 Sequence
23	53	100.0	83	8 ADQ91446	Adq91446 Amino aci
24	53	100.0	84	1 AAP60248	Aap60248 Human pro
25	53	100.0	85	2 AAR20711	Aar20711 Ser A13,

26	53	100.0	85	7 ADF17074	Adf17074 Human alb
27	53	100.0	86	1 AAF40829	Aap40829 Sequence
28	53	100.0	86	2 AAR20710	Aar20710 Proinsuli
29	53	100.0	86	2 AAR20699	Aar20699 Modified
30	53	100.0	86	2 AAR22327	Aar22327 Modified
31	53	100.0	86	2 AAR22330	Aar22330 Modified
32	53	100.0	86	2 AAR22331	Aar22331 Modified
33	53	100.0	86	2 AAR22332	Aar22332 Proinsuli
34	53	100.0	86	2 AAR20694	Aar20694 Modified
35	53	100.0	86	2 AAR22328	Aar22328 Modified
36	53	100.0	86	2 AAR20707	Aar20707 Modified
37	53	100.0	86	2 AAR22326	Aar22326 Modified
38	53	100.0	86	2 AAR22336	Aar22336 Modified
39	53	100.0	86	2 AAR20705	Aar20705 Proinsuli
40	53	100.0	86	2 AAR22329	Aar22329 Modified
41	53	100.0	86	2 AAR20693	Aar20693 Modified
42	53	100.0	86	2 AAR20712	Aar20712 Modified
43	53	100.0	86	2 AAR20704	Aar20704 Modified
44	53	100.0	86	2 AAR20692	Aar20692 Proinsuli
45	53	100.0	86	2 AAR22334	Aar22334 Proinsuli

ALIGNMENTS

RESULT 1
AAW06808
ID AAW06808 standard; peptide; 12 AA.

XX AAW06808;
AC
DT 16-MAR-1997 (first entry)
XX
DE Single chain insulin polypeptide bridging peptide.
XX
KW Single chain insulin; diabetes mellitus; gene therapy.
XX
OS Synthetic.
XX
PN EP741188-A2.
XX
PD 06-NOV-1996.
XX
PF 03-MAY-1996; 96EP-00303133.
XX
PR 05-MAY-1995; 95US-00435503.
XX
PR 05-MAY-1995; 95US-00435762.
XX
(ELIL) LILLY & CO ELI.

XX Chance RE, Dimarchi RD, Hoffmann JA, Long HB, Miller AR;
WPI; 1996-487391/49.
XX Single chain insulin polypeptide(s) - used for treating diabetes.
XX Disclosure; Page 16; 22pp; English.

XX Bridging peptides (AAW06808 and AAW00872) are utilised in novel single
chain peptides of formula b-BP-a (see also AAW06806-07), where b is the
insulin B-chain or its analogue, BP is a bridging peptide, and a is the
insulin A-chain or its analogue. These peptides show high insulin
bioactivity. Suitable bridging peptides should not contain sequences that
interact with the IGF-1 receptor, nor should they contain cleavage sites.
XX They are preferably derived from the C-peptide of human proinsulin so as
to minimise potential immunogenicity, and most preferably consist of
amino acids 54-65 of the C-peptide

XX Sequence 12 AA;

Query Match 100.0%; Score 53; DB 2; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.00059;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPLALEGSLOK 11
 Db 1 QPLALEGSLOK 11

RESULT 2
 ADI29027
 ID ADI29027 standard; peptide; 18 AA.
 AC ADI29027;
 XX
 DT 15-APR-2004 (first entry)
 XX
 DE - Proinsulin (73-90) peptide SEQ ID NO:23.
 XX
 KW MHC class II compound; MHC class II component; MHC class II alpha chain;
 KW MHC class II beta chain; peptide binding groove; spacerholder molecule;
 KW effector component; immune response; immune disorder; virucide;
 KW antibacterial; antiparasitic; cytostatic; immunosuppressive;
 KW gene therapy; viral infections; bacterial infection; parasitic infection;
 KW neoplastic disease; autoimmunity; toxicity.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 PN WO2004007528-A2.
 XX
 PD 22-JAN-2004.
 XX
 PF 11-JUL-2003; 2003WO-US021767.
 XX
 PR 12-JUL-2002; 2002US-0395494P.
 PR 22-JUL-2002; 2002US-0397593P.
 XX
 PA (DAND) DANA FARBER CANCER INST INC.
 XX
 PI Wucherpfenning KW, Seth N;
 XX
 DR WPI; 2004-122876/12.
 XX
 PT New MHC class II compound, useful for preparing a composition for
 PT treating immune disorders e.g. viral infections, bacterial infections,
 PT parasitic infections, neoplastic disease, autoimmunity or toxicity.
 PS Example 4; SEQ ID NO 23; 92pp; English.
 XX
 CC The present invention describes an isolated MHC class II compound (I)
 CC comprising: (a) an MHC class II component comprising at least a portion
 CC of an MHC class II alpha chain and at least a portion of an MHC class II
 CC beta chain, such that the MHC class II alpha chain and MHC class II beta
 CC chain form a peptide binding groove; (b) a spacerholder molecule; and (c)
 CC an effector component, where the effector component is linked to the MHC
 CC class II component. Also described: (1) a pharmaceutical composition
 CC comprising an MHC class II molecule and a carrier; (2) a method of
 CC producing an MHC class II compound; (3) a method of directly identifying
 CC an antigen-specific T cell; (4) a method of regulating an immune response
 CC in a subject; (5) a method of treating an immune disorder in a subject;
 CC (6) a method of regulating an immune response ex vivo in a subject; and
 CC (7) a method of treating an immune disorder ex vivo in a subject. (I) has
 CC virucide, antibacterial, antiparasitic, cytostatic and immunosuppressive
 CC activities, and can be used in gene therapy. The MHC class II compound
 CC (I) can be used for preparing a composition for treating immune
 CC disorders, e.g., viral infections, bacterial infections, parasitic
 CC infections, neoplastic disease, autoimmunity or toxicity. The present
 CC sequence represents a peptide which can be used to generate peptide-
 CC specific antibodies, which is used in an example from the present
 CC invention.
 XX
 SQ Sequence 18 AA;
 Query Match 100.0%; Score 53; DB 8; Length 18;
 Best Local Similarity 100.0%; Pred. No. 0.00094;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 QPLALEGSLOK 11
 Db 6 QPLALEGSLOK 16

RESULT 3
 AAR60842
 ID AAR60842 standard; peptide; 23 AA.
 XX
 AC AAR60842;
 XX
 DT 25-MAR-2003 (revised)
 DT 05-JUN-1995 (first entry)
 XX
 DE AOA-PepC COSM polyoxime component.
 XX
 KW Polyoxime; homopolyoxime; heteropolyoxime; peptide presentation;
 KW cell imaging; complementary orthogonal specifically active molecule;
 KW COSM; baseplate; proinsulin-C; insulin; immunogen.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Modified-site 1
 FT FT /label= OTHER
 FT FT /note= "aminoxyacetyl-glutamic acid"
 XX
 PN WO9425071-A1.
 XX
 PD 10-NOV-1994.
 XX
 PF 05-MAY-1994; 94WO-IB000093.
 XX
 PR 05-MAY-1993; 93US-00057594.
 PR 31-AUG-1993; 93US-00105904.
 PR 31-AUG-1993; 93US-00114877.
 XX
 PA (ROSE/) ROSE K.
 PA (OFFO/) OFFORD R E.
 XX
 PI Rose K, Offord RE;
 XX
 DR WPI; 1994-357918/44.
 XX
 PT Homo- and hetero-polyoxime compounds and their preparation - used for
 PT peptide presentation to antibodies and in cell imaging etc.
 XX
 PS Disclosure; Page 53; 85pp; English.
 XX
 CC Peptides given in AAR60833-62 are used as baseplates and COSMs for the
 CC preparation of polyoximes having varying spacing, charge, lipophilicity,
 CC valency, conformational restraints, solubility and other physical and
 CC biological properties. The AOA PepC COSM AAR60842 is based on the human
 CC proinsulin C peptide sequence. (Updated on 25-MAR-2003 to correct PN
 CC field.)
 XX
 SQ Sequence 23 AA;
 Query Match 100.0%; Score 53; DB 2; Length 23;
 Best Local Similarity 100.0%; Pred. No. 0.0012;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPLALEGSLOK 11
 Db 12 QPLALEGSLOK 22

RESULT 4
 AAW81858
 ID AAW81858 standard; peptide; 27 AA.
 XX

Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 4, Appli
Sequence 5, Appli
Sequence 125, App
Sequence 3, Appli
Sequence 9, Appli
Sequence 6, Appli
Sequence 125, App
Sequence 1, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 44, Appl
Sequence 44, Appl
Sequence 44, Appl
Sequence 20, Appl
Sequence 44, Appl
Sequence 44, Appl
Sequence 44, Appl
Sequence 44, Appl
Sequence 3, Appli
Sequence 44, Appl
Sequence 63, Appl
Sequence 62, Appl
Sequence 1, Appli
Sequence 2, Appli
Sequence 1, Appli
Sequence 15, Appl

86 15 US-10-444-262-2
86 15 US-10-444-649-2
86 15 US-10-444-701-2
96 9 US-09-947-563-4
96 9 US-09-947-563-5
110 9 US-09-205-658-125
110 9 US-09-815-229-3
110 9 US-09-804-409A-9
110 10 US-09-969-748C-6
110 10 US-09-963-693-125
110 14 US-10-038-686-1
110 14 US-10-328-813-2
110 14 US-10-383-285-2
110 14 US-10-345-563-2
110 15 US-10-321-717-2
110 15 US-10-411-037-44
110 15 US-10-411-026-44
110 15 US-10-410-962-44
110 15 US-10-411-049-44
110 15 US-10-700-725-20
110 16 US-10-410-930-44
110 16 US-10-410-997-44
110 16 US-10-411-012-44
110 16 US-10-287-994-44
110 16 US-10-740-098-3
110 16 US-10-410-913-44
117 9 US-09-280-030-63
130 9 US-09-369-439-1
31 9 US-09-878-380-2
31 15 US-10-430-752A-1
31 15 US-10-430-752A-15

ALIGNMENTS

RESULT 1
US-10-617-568-23
; Sequence 23, Application US/10617568
; Publication NO. US20040197862A1
; GENERAL INFORMATION:
; APPLICANT: Dana-Farber Cancer Institute, Inc. et al.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR THE
; GENERATION OF MHC CLASS II COMPOUNDS BY
; TITLE OF INVENTION: PEPTIDE EXCHANGE
; FILE REFERENCE: DFN-044
; CURRENT APPLICATION NUMBER: US/10/617,568
; CURRENT FILING DATE: 2003-07-11
; PRIOR APPLICATION NUMBER: 60/395494
; PRIOR FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 60/397893
; PRIOR FILING DATE: 2002-07-22
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-617-568-23

Query Match 100.0%; Score 11; DB 17; Length 18;
Best Local Similarity 100.0%; Pred. No. 0.00037;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 QPLALEGSLQK 11
Db 6 QPLALEGSLQK 16
RESULT 2
US-10-641-834-24
; Sequence 24, Application US/10641834

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.
OM protein - protein search, using sw model
Run on: October 28, 2004, 11:39:31 ; Search time 21.7349 Seconds
(without alignments)
164.085 Million cell updates/sec

Title: US-10-783-095-9
Perfect score: 11
Sequence: 1 QPLALEGSLQK 11
Scoring table: OLIGO
Gapop 60.0, Gapext 60.0
Searched: 1370721 seqs, 324215800 residues
Word size : 0
Total number of hits satisfying chosen parameters: 1370721
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Listing first 45 summaries
Database : Published Applications AA:
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	11	100.0	18	17	US-10-617-568-23
2	11	100.0	33	17	US-10-641-834-24
3	11	100.0	35	9	US-09-815-229-4
4	11	100.0	35	9	US-09-947-563-3
5	11	100.0	35	16	US-10-740-098-4
6	11	100.0	54	9	US-09-815-229-13
7	11	100.0	54	16	US-10-740-098-13
8	11	100.0	86	9	US-09-878-380-1
9	11	100.0	86	10	US-09-858-935B-4
10	11	100.0	86	13	US-10-028-410-2
11	11	100.0	86	13	US-10-054-873-4
12	11	100.0	86	14	US-10-444-326-2
13	11	100.0	86	14	US-10-271-869-4

```
; Publication No. US20040209830A1
; GENERAL INFORMATION:
; APPLICANT: Russell, Stephen James
; APPLICANT: Peng, Kah Whye
; TITLE OF INVENTION: SYSTEM FOR MONITORING THE EXPRESSION OF
; FILE OF INVENTION: TRANSGENES
; FILE REFERENCE: 07039-292002
; CURRENT APPLICATION NUMBER: US/10/641,834
; CURRENT FILING DATE: 2003-05-18
; PRIOR APPLICATION NUMBER: 09/639,667
; PRIOR FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: 60/149,168
; PRIOR FILING DATE: 1999-08-17
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 33
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: C-peptide sequence
US-10-641-834-24

Query Match 100.0%; Score 11; DB 17; Length 33;
Best Local Similarity 100.0%; Pred. No. 0.00064;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QPLALEGSLQK 11
Db 22 QPLALEGSLQK 32

RESULT 3
US-09-815-229-4
; Sequence 4, Application US/09815229
; Patent No. US20020058614A1
; GENERAL INFORMATION:
; APPLICANT: Filvaroff, Ellen H.
; APPLICANT: Okumu, Franklin W.
; TITLE OF INVENTION: USE OF INSULIN FOR THE TREATMENT OF CARTILAGENOUS DISORDERS
; FILE REFERENCE: P178631US
; CURRENT APPLICATION NUMBER: US/09/815,229
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/192,103
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 4
; LENGTH: 35
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-815-229-4

Query Match 100.0%; Score 11; DB 9; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.00067;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QPLALEGSLQK 11
Db 24 QPLALEGSLQK 34

RESULT 4
US-09-947-563-3
; Sequence 3, Application US/09947563
; Patent No. US20020156234A1
; GENERAL INFORMATION:
; APPLICANT: Rubroder, Franz-Josef
; APPLICANT: Keller, Reinhold
; TITLE OF INVENTION: Improved process for obtaining
; INSULIN precursors having correctly bonded cystine bridges
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farrah, Garrett &
```

```
; Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/09/947,563
; APPLICATION NUMBER: 09/134,836
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Leslie McDonnell
; REGISTRATION NUMBER: 34,872
; REFERENCE/DOCKET NUMBER: 02481.1600-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..35
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-947-563-3

Query Match 100.0%; Score 11; DB 9; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.00067;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QPLALEGSLQK 11
Db 24 QPLALEGSLQK 34

RESULT 5
US-10-740-098-4
; Sequence 4, Application US/10740098
; Publication No. US20040138101A1
; GENERAL INFORMATION:
; APPLICANT: Filvaroff, Ellen H.
; APPLICANT: Okumu, Franklin W.
; TITLE OF INVENTION: USE OF INSULIN FOR THE TREATMENT OF CARTILAGENOUS DISORDERS
; FILE REFERENCE: P178631US
; CURRENT APPLICATION NUMBER: US/10/740,098
; CURRENT FILING DATE: 2003-12-17
; PRIOR APPLICATION NUMBER: US/09/815,229
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/192,103
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 4
; LENGTH: 35
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-740-098-4

Query Match 100.0%; Score 11; DB 16; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.00067;
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 28, 2004, 16:28:53 ; Search time 721.759 Seconds
(without alignments)
720.721 Million cell updates/sec

Title: US-10-783-095-9

Perfect score: 11

Sequence: 1 QPLALEGSLQK 11

Scoring table: OLIGO

Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4526729 seqs, 23644849745 residues

Word size: 1

Total number of hits satisfying chosen parameters: 9041963

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters: -DEV=xlp

-O=/cgn2.1/USPTO.spool_p/US10783095/runat_26102004.185454.25032/app_query.fasta_1.796
-DB=GenEmbl -QFMT=fastcap -SUFFIX=oli.rge -MINMATCH=0.1 -LOPCU=0 -LOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10783095 @CNC 1 1.6148 @runat_26102004.185454.25032 -NCPU=6 -ICPU=3
-NO_NMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=6 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hgt:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_scs:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	11	100.0	68	AR077834	Sequence
2	11	100.0	87	AX107018	Sequence
3	11	100.0	105	12 XXHSIN	V01461 Synthetic g
4	11	100.0	216	6 AX179492	AX179492 Sequence

5	11	100.0	252	6	I01821	101821 Sequence 6
6	11	100.0	258	6	AR077833	AR077833 Sequence
7	11	100.0	251	6	A48809	A48809 Sequence 1
8	11	100.0	261	6	A48810	A48810 Sequence 2
9	11	100.0	261	6	A61947	A61947 Sequence 27
10	11	100.0	269	6	AR069914	AR069914 Sequence
11	11	100.0	271	6	B43909	B43909 Novel vecto
12	11	100.0	272	6	I02917	I02917 Sequence 6
13	11	100.0	273	6	E00074	E00074 DNA sequenc
14	11	100.0	273	6	I00910	I00910 Sequence 2
15	11	100.0	275	6	I14159	I14159 Sequence 12
16	11	100.0	275	6	I14160	I14160 Sequence 14
17	11	100.0	276	6	I14161	I14161 Sequence 16
18	11	100.0	277	12	ARHSINS	V00082 Artificial
19	11	100.0	279	6	E43908	E43908 Novel vecto
20	11	100.0	286	12	SYNINGSGA	M12913 Synthetic h
21	11	100.0	289	6	I03833	I03833 Sequence 35
22	11	100.0	290	6	A07328	A07328 Synthetic D
23	11	100.0	290	6	A07329	A07329 Synthetic D
24	11	100.0	290	6	A11939	A11939 Nucleotide
25	11	100.0	290	6	A11940	A11940 Nucleotide
26	11	100.0	290	6	I12514	I12514 Sequence 41
27	11	100.0	290	6	I12515	I12515 Sequence 42
28	11	100.0	298	6	I13722	I13722 Sequence 28
29	11	100.0	330	6	AR275652	AR275652 Sequence
30	11	100.0	333	6	BD224077	BD224077 Method fo
31	11	100.0	333	9	BT006808	BT006808 Homo sapi
32	11	100.0	333	12	BT007778	BT007778 Synthetic
33	11	100.0	342	6	I00596	I00596 Sequence 4
34	11	100.0	342	6	I02324	I02324 Sequence 4
35	11	100.0	342	6	I03602	I03602 Sequence 4
36	11	100.0	342	6	I07905	I07905 Sequence 15
37	11	100.0	346	6	E00054	E00054 DNA encodin
38	11	100.0	346	6	E00191	E00191 DNA coding
39	11	100.0	346	6	E00575	E00575 Synthetic D
40	11	100.0	351	6	E39198	E39198 DNA encodin
41	11	100.0	351	6	AR274665	AR274665 Sequence
42	11	100.0	351	12	SYNINGS	J02547 Human (synt
43	11	100.0	359	6	AR102882	AR102882 Sequence
44	11	100.0	359	6	AR108061	AR108061 Sequence
45	11	100.0	359	6	AR134709	AR134709 Sequence

ALIGNMENTS

RESULT 1
AR077834
LOCUS AR077834 68 bp DNA linear PAT 31-AUG-2000
DEFINITION Sequence 18 from patent US 5962267.
ACCESSION AR077834
VERSION AR077834.1 GI:10004580
KEYWORDS .
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68)
AUTHORS Shin,H.-C., Chang,S.-G., Kim,D.-Y. and Kim,C.-S.
TITLE Proinsulin derivative and process for producing human insulin
JOURNAL Patent: US 5962267-A 18 05-OCT-1999;
FEATURES
Location/Qualifiers
source 1..68
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Alignment Scores:
Pred. No.: 0.000267 Length: 68
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-783-095-9 (1-11) x AR077834 (1-68)

QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnLys 11
 DB 7 CAGCCGTTGGCGCTGGAAGGTTCCCTGCAGAAA 39

RESULT 2
 AX107018 87 bp DNA linear PAT 30-APR-2001
 LOCUS
 DEFINITION Sequence 1 from Patent WO0125278.
 ACCESSION AX107018
 VERSION AX107018.1 GI:13922551
 KEYWORDS
 SOURCE synthetic construct
 ORGANISM synthetic construct
 artificial sequences.

REFERENCE 1
 AUTHORS Habermann, P., Ertl, J., Meiwes, J. and Seipke, G.
 TITLE C-peptide for the improved production of insulin and insulin analogues
 JOURNAL Patent: WO 0125278-A 1 12-APR-2001;
 FEATURES Aventis Pharma Deutschland GmbH (DE)
 source Location/Qualifiers
 1..87
 /organism="synthetic construct"
 /mol_type="unassigned DNA"
 /db_xref="taxon:32630"
 /note="DNA kodierend fur C-PEPTIDEVARIANTE"

ORIGIN
 Alignment Scores:
 Pred. No.: 0.000329 Length: 87
 Score: 11.00 Matches: 11
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-10-783-095-9 (1-11) x AX107018 (1-87)
 QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnLys 11
 DB 52 CAGCCGTTGGCGCTGGAAGGTTCCCTGCAGAA 84

RESULT 3
 XXHSIN 105 bp DNA linear SYN 04-AUG-1995
 LOCUS
 DEFINITION Synthetic gene for human insulin C-chain.
 ACCESSION V01461 J02546
 VERSION V01461.1 GI:58303
 KEYWORDS artificial gene; gene synthesis.
 SOURCE synthetic construct
 ORGANISM synthetic construct
 artificial sequences.

REFERENCE 1 (bases 1 to 105)
 AUTHORS Hsiung, H.M., Sung, W.L., Brousseau, R., Wu, R. and Narang, S.A.
 TITLE Synthesis of the human insulin gene. Part III. Chemical synthesis of 5'-phosphonoester group containing deoxyribonucleotides by the modified phosphotriester method. Its application in the synthesis of seventeen fragments constituting human insulin C-chain DNA

Nucleic Acids Res. 8 (23), 5753-5765 (1980)

JOURNAL Nucleic Acids Res. 8 (23), 5753-5765 (1980)
 MEDLINE 81124297
 PUBMED 7008029
 FEATURES Location/Qualifiers
 source 1..105
 /organism="synthetic construct"
 /mol_type="other DNA"
 /db_xref="taxon:32630"
 <1..>105
 /note="(amino acids 31 to 65)"
 /codon_start=1
 /transl_table=11

CDS

/product="insulin C chain"
 /protein_id="CAA24707.1"
 /db_xref="GI:929915"
 /translation="RREARDLQVGQVELGGFGAGSLQPLALEGSLQXR"

ORIGIN

Alignment Scores:
 Pred. No.: 0.000387 Length: 105
 Score: 11.00 Matches: 11
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 12 Gaps: 0

US-10-783-095-9 (1-11) x XXHSIN (1-105)

QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnLys 11
 DB 70 CAACCTTGGCTCGAGGATCACTTCAAAAG 102

RESULT 4
 AX179492 216 bp DNA linear PAT 31-AUG-2001
 LOCUS
 DEFINITION Sequence 87 from Patent WO0131037.
 ACCESSION AX179492
 VERSION AX179492.1 GI:14599120
 KEYWORDS
 SOURCE synthetic construct
 ORGANISM synthetic construct
 artificial sequences.

REFERENCE 1
 AUTHORS Ben-Nun, A., Kerlero de Rosbo, N. and Sappler, G.P.
 TITLE Synthetic human genes and polypeptides and their use in the treatment of autoimmune diseases
 JOURNAL Patent: WO 0131037-A 87 03-MAY-2001;
 YEDA RESEARCH AND DEVELOPMENT CO. LTD. (IL)
 FEATURES Location/Qualifiers
 1..216
 /organism="synthetic construct"
 /mol_type="unassigned DNA"
 /db_xref="taxon:32630"
 1..207
 /note="unnamed protein product; synthetic construct"

source

/codon_start=1
 /transl_table=11
 /protein_id="CAC43900.1"
 /db_xref="GI:14599121"
 /translation="MPRELLPLALLALWGPDPAAASHLVEALYLVSGRGFFYTPKTR
 TEAAGAGSLQPLALEGSLQKIPR"

ORIGIN

Alignment Scores:
 Pred. No.: 0.000721 Length: 216
 Score: 11.00 Matches: 11
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-10-783-095-9 (1-11) x AX179492 (1-216)

QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnLys 11
 DB 163 CACCCGTTGGCCTGGAGGGTCCCTGCAAAAG 195

RESULT 5

101821
 LOCUS
 DEFINITION Sequence 6 from Patent US 4801536.
 ACCESSION 101821
 VERSION 101821.1 GI:269797
 KEYWORDS
 SOURCE Unknown.

I01821

Sequence 6 from Patent US 4801536.

linear

PAT 21-MAY-1993

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 28, 2004, 16:24:37 ; Search time 143.265 Seconds
(without alignments)
403.055 Million cell updates/sec

Title: US-10-783-095-9

Perfect score: 11
Sequence: 1 QPLAEGSLQK 11

Scoring table:

OLIGO	60.0	Xgapext	60.0
Ygapop	60.0	Ygapext	60.0
Fgapop	6.0	Fgapext	7.0
Delcp	6.0	Delext	7.0

Searched: 4134886 seqs, 2624710521 residues

Word size: 1

Total number of hits satisfying chosen parameters: 8253356

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n.model -DEV=xlp
-Q/cgn2_1/USPTO spo0l_p/US10783095/runat_26102004_185454_25024/app_query.fasta_1.796
-DB=N Geneseq 23Sep04 -QFMT=fastap -SUFFIX=oli.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10783095@CN 1_1_1241@runat_26102004_185454_25024 -NCFU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-TIMECUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N Geneseq 23Sep04:*

1:	Geneseqn1980s:*
2:	Geneseqn1990s:*
3:	Geneseqn2000s:*
4:	Geneseqn2001as:*
5:	Geneseqn2001bs:*
6:	Geneseqn2002as:*
7:	Geneseqn2002bs:*
8:	Geneseqn2003as:*
9:	Geneseqn2003bs:*
10:	Geneseqn2003cs:*
11:	Geneseqn2003ds:*
12:	Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	11	100.0	66	2 AAX86082	Aax86082 PCR prime
2	11	100.0	87	2 AAF61522	Aaf61522 Human pro
3	11	100.0	159	2 AAT45977	Aat45977 Oligonuc1
4	11	100.0	189	2 AAT45973	Aat45973 DNA encod
5	11	100.0	201	2 AAT45974	Aat45974 DNA encod
6	11	100.0	216	5 AAS04793	Aas04793 Synthetic

7	11	100.0	252	1 AAN70988	Aan70988 Sequence
8	11	100.0	258	2 AAT37134	Aat37134 Proinsuli
9	11	100.0	258	10 ADC51568	Adc51568 Human pro
10	11	100.0	258	10 ADF16306	Adf16306 Human alb
11	11	100.0	258	10 ADH21708	Adh21708 Human lon
12	11	100.0	260	12 ADH04119	Adh04119 Proinsuli
13	11	100.0	261	2 AAT17831	Aat17831 Optimised
14	11	100.0	261	2 AAT17830	Aat17830 Human ins
15	11	100.0	264	2 AAT45975	Aat45975 Gene for
16	11	100.0	268	2 AAO43935	Aa43935 MetTyr hu
17	11	100.0	269	2 AAT80401	Aat80401 Proinsuli
18	11	100.0	272	1 AAN20041	Aan20041 Human pro
19	11	100.0	272	1 AAN60872	Aan60872 Sequence
20	11	100.0	275	2 AAO37003	Aa37003 Proinsuli
21	11	100.0	278	2 AAO32256	Aa32256 Synthetic
22	11	100.0	279	2 AAO83195	Aa83195 Proinsuli
23	11	100.0	281	2 AAO05663	Aa05663 Human pro
24	11	100.0	281	2 AAO38310	Aa38310 hpl gene.
25	11	100.0	283	2 AAO20543	Aa20543 Encodes p
26	11	100.0	287	2 AAT04874	Aat04874 Synthetic
27	11	100.0	289	12 ADL24436	Adl24436 Modified
28	11	100.0	290	12 ADL24441	Adl24441 Modified
29	11	100.0	293	1 AAN90279	Aan90279 Synthetic
30	11	100.0	298	2 AAO27123	Aa27123 DNA encod
31	11	100.0	298	2 AAO99025	Aa99025 Chelating
32	11	100.0	304	2 AAO15230	Aa15230 NcoI-Sali
33	11	100.0	327	6 AAN85338	Aan85338 Oligonuc1
34	11	100.0	330	9 ADA09217	Ada09217 Human mRN
35	11	100.0	330	12 ADJ25929	Adj25929 Human pre
36	11	100.0	333	3 AA251476	Aa251476 Human ins
37	11	100.0	342	1 AAN40179	Aan40179 Sequence
38	11	100.0	342	1 AAN50152	Aan50152 Sequence
39	11	100.0	342	1 AAN50082	Aan50082 Proinsuli
40	11	100.0	351	3 AAZ59208	Aaz59208 MWPeP-MWP
41	11	100.0	359	2 AAT75193	Aat75193 Human ins
42	11	100.0	359	2 AAT75652	Aat75652 Human ins
43	11	100.0	359	2 AAC55717	Aac55717 Mutant hu
44	11	100.0	359	4 AAF58802	Aaf58802 Human ins
45	11	100.0	360	5 AAS04761	Aas04761 Synthetic

ALIGNMENTS

RESULT 1
ID AAX86082/c AAX86082 standard; DNA; 66 BP.
XX AC AAX86082;
XX DT 15-SEP-1999 (first entry)
XX DE PCR primer used to produce chimeric genes of the invention.
XX KW Chimeric fusion protein; immunodominant epitope; glutamate decarboxylase;
KW GAD; insulin; insulin chain B; human T cell response; GAD65 peptide;
KW Type I diabetes; insulin dependent diabetes mellitus; IDDM;
KW Stiff Man syndrome; transplant; pancreatic cell; PCR primer; ss.
XX OS Synthetic.
XX OS Homo sapiens.
XX FN WO932136-A1.
XX PD 01-JUL-1999.
XX PF 23-DEC-1998; 98WO-US027408.
XX PR 23-DEC-1997; 97US-0068648P.
XX PA (ALEX-) ALEXION PHARM INC.
XX PA (MUEL/) MUELLER J P.
XX PA (MATI/) MATIS L A.

PI Mueller JP, Matis LA, Wang Y;
 XX WPI; 1999-405113/34.
 XX
 PT Chimeric fusion proteins comprising immunodominant epitopes of glutamate
 PT decarboxylase and insulin, used to treat patients that have or are
 PT predicted to be at risk of developing Type I diabetes.
 XX
 PS Example; Page 62; 67pp; English.
 XX
 CC The specification describes chimeric fusion proteins comprising
 CC immunodominant epitopes of glutamate decarboxylase (GAD) and insulin. The
 CC chimeric fusion proteins comprises insulin chain B and individual peptide
 CC moieties consisting of at least one GAD65 peptide capable of eliciting a
 CC human T cell response, where the insulin chain B and GAD65 peptide are
 CC covalently linked and the chimeric fusion protein is capable of eliciting
 CC a human T cell response to insulin chain B and to each of the at least
 CC one GAD65 peptides. The chimeric fusion proteins are used to treat
 CC patients that are predicted to be at risk of developing type I diabetes
 CC and those suffering from type I diabetes. The fusion proteins can also be
 CC used to diagnose patients suffering from insulin dependent diabetes
 CC mellitus (IDDM) and/or Stiff Man syndrome. This is also useful for
 CC determining the suitability of patients as recipients of transplants of
 CC pancreatic cells or tissues. The present sequence represents a primer
 CC used to construct genes encoding the chimeric fusion proteins of the
 CC invention
 XX
 SQ Sequence 66 BP; 10 A; 27 C; 21 G; 8 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 0.00155 Length: 66
 Score: 11.00 Matches: 11
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 2 Gaps: 0
 US-10-783-095-9 (1-11) x AAX86082 (1-66)
 QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnLys 11
 DB 42 CAGCCCTTGGCCCTGGAGGGTCCCTGCAGAAG 10
 RESULT 2
 ID AAF61522 standard; DNA; 87 BP.
 XX
 AC AAF61522;
 XX
 DT 04-JUL-2001 (first entry)
 XX
 DE Human proinsulin C-peptide encoding DNA.
 XX
 KW Insulin; proinsulin; C-peptide; human; B-chain; A-chain; diabetes; ds.
 XX
 OS Homo sapiens.
 XX
 PN DE19947456-A1.
 XX
 PD 05-APR-2001.
 XX
 PF 02-OCT-1999; 99DE-01047456.
 XX
 PR 02-OCT-1999; 99DE-01047456.
 XX
 PA (AVET) AVENTIS PHARMA DEUT GMBH.
 XX
 PI Habermann P, Ertl J, Meives J, Seipke G;
 XX WPI; 2001-267042/28.
 DR P-PSDB; AAB70847.
 XX
 PT New synthetic derivatives of the C-peptide of proinsulin, useful in the

PT preparation of human insulin or insulin analogs in high yield.
 XX
 PS Disclosure; Page 7; 12pp; German.
 XX
 CC This invention describes novel synthetic derivatives (I) of the C-peptide
 CC of proinsulin. Such novel derivatives of human insulin or insulin analogs
 CC represented by (I) are of formula Fus-B(1-30)-RDVP-Yn-A(1-21). Fus =
 CC optionally present fusion component of arbitrary sequence; B(1-30) = the
 CC B-chain of human insulin; Y = amino acid chain having a basic amino acid
 CC at the C-terminal; n = 2-50 (defining the chain length of Y), and A(1-21)
 CC = the A-chain of human insulin; the A- and/or B-chain(s) may be modified
 CC by amino acid exchanges, additions and/or deletions. (I) are
 CC intermediates for human insulin or its analogs, which are used in the
 CC treatment of diabetes. A claimed method of preparing human insulin (or
 CC analogs) involves preparing (I), folding (I) so that disulfide bonds as
 CC in human insulin can form, enzymatically removing the RDVP-Yn part and
 CC Fus (if present) and purifying the product. Insulin (or analogs) can be
 CC prepared in increased yields via (I). Specifically the expression yield
 CC can be increased around 20% compared with that obtained using plasmid
 CC pINR90d and almost 5-fold compared with that obtained using plasmid
 CC pINR302d. The control of enzymatic processing is also improved. The
 CC folding rate is comparable with that obtained using simian proinsulin
 CC encoded by pINR90d. This sequence encodes the human proinsulin C-peptide
 CC which is described in the method of the invention
 XX
 SQ Sequence 87 BP; 10 A; 27 C; 37 G; 13 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 0.00201 Length: 87
 Score: 11.00 Matches: 11
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 4 Gaps: 0
 US-10-783-095-9 (1-11) x AAF61522 (1-87)
 QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnLys 11
 DB 52 CAGCCCTTGGCCCTGGAGGGTCCCTGCAGAAG 84
 RESULT 3
 ID AAT45977 standard; cDNA; 159 BP.
 XX
 AC AAT45977;
 XX
 DT 16-MAR-1997 (first entry)
 XX
 DE Oligonucleotide for 3' end of single chain insulin.
 XX
 KW Single chain insulin; diabetes mellitus; gene therapy; ss.
 XX
 OS Synthetic.
 XX
 PN EP741188-A2.
 XX
 PD 06-NOV-1996.
 XX
 PF 03-MAY-1996; 96EP-00303133.
 XX
 PR 05-MAY-1995; 95US-00435503.
 XX
 PR 05-MAY-1995; 95US-00435762.
 XX
 PA (ELIL) LILLY & CO ELI.
 XX
 PI Chance RE, Dimarchi RD, Hoffmann JA, Long HB, Miller AR;
 XX WPI; 1996-487391/49.
 DR
 XX Single chain insulin polypeptide(s) - used for treating diabetes.
 XX
 PS Example 7; Page 15; 22pp; English.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 28, 2004, 20:02:28 ; Search time 32.6024 Seconds
(without alignments)
239.819 Million cell updates/sec

Title: US-10-783-095-9

Perfect score: 11

Sequence: 1 QPLAEGSLQK 11

Scoring table:

OLIGO XGapop 60.0 , Xgapext 60.0
YGapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 824507 seqs, 355394441 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1641062

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-O=/cgn2_1/USPTC spool_p/US10783095/runat_26102004_185455_25065/app_query.fasta_1.796
-DB=Issued patents NA -QFMT=fastap -SUFFIX=oli.rni -MINMATCH=0.1 -LQOPCI=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THE SCORE=quality -THE MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZES=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10783095@cgn_1_1_213@runat_26102004_185455_25065 -NCPUS=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Issued Patents NA.*

1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTCUS.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	11	100.0	68	2	US-08-600-783-18
2	11	100.0	87	4	US-09-676-787-1
3	11	100.0	258	2	US-08-600-783-17
4	11	100.0	269	2	US-08-673-312-10
5	11	100.0	275	1	US-07-918-953-12
6	11	100.0	275	1	US-07-918-953-14
7	11	100.0	275	1	US-08-081-661-12
8	11	100.0	275	1	US-08-081-661-14
9	11	100.0	276	1	US-07-918-953-16
10	11	100.0	276	1	US-08-081-661-16
c 11	11	100.0	277	1	US-07-715-183C-3
12	11	100.0	278	1	US-07-715-183C-2

13	11	100.0	281	1	US-07-764-655D-12
c 14	11	100.0	281	1	US-07-764-655D-13
15	11	100.0	281	6	5514646-1
16	11	100.0	290	1	US-08-030-731A-41
c 17	11	100.0	290	1	US-08-030-731A-42
18	11	100.0	298	1	US-07-826-928A-28
19	11	100.0	304	1	US-07-596-551B-12
20	11	100.0	330	4	US-08-472-701-1
21	11	100.0	330	5	PCT-US95-08596-1
22	11	100.0	351	4	US-09-280-030-49
23	11	100.0	359	3	US-08-589-028-3
24	11	100.0	359	3	US-08-784-582-3
25	11	100.0	359	3	US-08-785-271-3
26	11	100.0	390	4	US-09-280-030-48
27	11	100.0	416	3	US-08-945-140-5
28	11	100.0	450	4	US-09-185-852-1
29	11	100.0	450	4	US-09-323-738-1
30	11	100.0	510	1	US-07-918-953-7
31	11	100.0	510	1	US-08-081-661-7
32	11	100.0	515	3	US-08-589-028-1
33	11	100.0	515	3	US-08-784-582-1
34	11	100.0	515	3	US-08-785-271-1
35	11	100.0	2499	4	US-09-775-508C-7
36	11	100.0	4952	4	US-09-015-399-10
c 37	10	90.9	51	6	5514646-42
38	10	90.9	51	6	5514646-43
39	10	90.9	68	4	US-09-485-286-7
c 40	10	90.9	74	4	US-09-485-286-9
41	10	90.9	4646	4	US-09-485-286-14
42	8	72.7	99	1	US-07-918-953-3
43	8	72.7	99	1	US-08-081-661-3
44	8	72.7	598	3	US-09-537-696-11
45	8	72.7	633	3	US-09-537-696-12

ALIGNMENTS

RESULT 1

US-08-600-783-18
; Sequence 18, Application US/08600783
; Patent No. 5962267
; GENERAL INFORMATION:
; APPLICANT: SHIN, Hang Cheol
; APPLICANT: CHANG, Seung Gu
; APPLICANT: KIM, Dae Young
; APPLICANT: KIM, Chong Suh
; TITLE OF INVENTION: Proinsulin Derivative and Process
; TITLE OF INVENTION: for Producing Human Insulin
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SHIN, Hang Cheol
; STREET: Ssangma-Hanshin Apt. 102-1206,
; STREET: #245 Cholsan-dong
; CITY: Kwangmyung-shi
; STATE: Kyungki-do
; COUNTRY: Republic of Korea
; ZIP: 423-030
; ADDRESSEE: CHANG, Seung Gu
; STREET: Hyundai Apt. 71-203, Apkujong-dong,
; STREET: Kangnam-ku
; CITY: Seoul
; STATE: Republic of Korea
; COUNTRY: Republic of Korea
; ZIP: 135-110
; ADDRESSEE: KIM, Dae Young
; STREET: Sosa Jukong Apt. 108-202, Sosa Bon-dong,
; STREET: Sosa-ku
; CITY: Bucheon-shi
; STATE: Kyungki-do
; COUNTRY: Republic of Korea
; ZIP: 422-230
; ADDRESSEE: KIM, Chong Suh
; STREET: Garden Heights Apt. 202-801, #100,
; STREET: Garden Heights Apt. 202-801, #100,

Sequence 12, Appl
Sequence 13, Appl
Patent No. 5514646
Sequence 41, Appl
Sequence 42, Appl
Sequence 28, Appl
Sequence 12, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 49, Appl
Sequence 3, Appl
Sequence 3, Appl
Sequence 3, Appl
Sequence 48, Appl
Sequence 5, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 7, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 10, Appl
Patent No. 5514646
Patent No. 5514646
Sequence 7, Appl
Sequence 9, Appl
Sequence 14, Appl
Sequence 3, Appl
Sequence 3, Appl
Sequence 11, Appl
Sequence 12, Appl

STREET: Hwangkeum-dong, Soosung-ku
 CITY: Taegu
 STATE: Taegu
 COUNTRY: Republic of Korea
 ZIP: 706-040
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy Disk, 3.5 inch, 1.44MB storage
 COMPUTER: IBM PC/AT
 OPERATING SYSTEM: MS-DOS
 SOFTWARE: Word Perfect 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/600,783
 FILING DATE:
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: KR 95-2751
 FILING DATE: 15-FEB-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Shahan Islam
 REGISTRATION NUMBER: 32,507
 REFERENCE/DOCKET NUMBER:
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 278-1000
 TELEFAX: (212) 953-7249
 INFORMATION FOR SEQ ID NO: 18:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 68 bases
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: oligonucleotide
 ANTI-SENSE: yes
 US-08-600-783-18

Alignment Scores:
 Pred. No.: 0.000207 Length: 68
 Score: 11.00 Matches: 11
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 2 Gaps: 0

US-10-783-095-9 (1-11) x US-08-600-783-18 (1-68)

QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnlys 11
 DB 7 CAGCCGTTGGCGCTGGAGGTTCCCTGCAGAA 39

RESULT 2

US-09-676-787-1
 Sequence 1, Application US/09676787
 Patent No. 6534288
 GENERAL INFORMATION:
 APPLICANT: HABERMANN, Paul
 APPLICANT: ERTL, Johann
 APPLICANT: MEIWES, Johannes
 APPLICANT: SEIPEKE, Gerhard
 TITLE OF INVENTION: C PEPTIDE FOR IMPROVED PREPARATION OF INSULIN AND INSULIN ANALOGS
 FILE REFERENCE: 38005-0110
 CURRENT APPLICATION NUMBER: US/09/676,787
 CURRENT FILING DATE: 2000-10-02
 PRIOR APPLICATION NUMBER: DE 199 47 456.7
 PRIOR FILING DATE: 1999-10-02
 NUMBER OF SEQ ID NOS: 13
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 1
 LENGTH: 87
 TYPE: DNA
 ORGANISM: Homo sapiens
 US-09-676-787-1

Alignment Scores:

Pred. No.: 0.000262 Length: 87

Score: 11.00 Matches: 11
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 4 Gaps: 0

US-10-783-095-9 (1-11) x US-09-676-787-1 (1-87)

QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnlys 11
 DB 52 CAGCCGTTGGCGCTGGAGGTTCCCTGCAGAA 84

RESULT 3

US-08-600-783-17
 Sequence 17, Application US/08600783
 Patent No. 5962267
 GENERAL INFORMATION:
 APPLICANT: SHIN, Hang Cheol
 APPLICANT: CHANG, Seung Gu
 APPLICANT: KIM, Dae Young
 APPLICANT: KIM, Chong Suh
 TITLE OF INVENTION: Proinsulin Derivative and Process
 TITLE OF INVENTION: for Producing Human Insulin
 NUMBER OF SEQUENCES: 36
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SHIN, Hang Cheol
 STREET: Ssangma-Hanshin Apt. 102-1206,
 STREET: #245 Cholsan-dong
 CITY: Kwangmyung-shi
 STATE: Kyungki-do
 COUNTRY: Republic of Korea
 ZIP: 423-030
 ADDRESSEE: CHANG, Seung Gu
 STREET: Hyundai Apt. 71-203, Apkujong-dong,
 STREET: Kangnam-ku
 CITY: Seoul
 STATE: Seoul
 COUNTRY: Republic of Korea
 ZIP: 135-110
 ADDRESSEE: KIM, Dae Young
 STREET: Sosa Jukong Apt. 108-202, Sosa Bon-dong,
 STREET: Sosa-ku
 CITY: Bucheon-shi
 STATE: Kyungki-do
 COUNTRY: Republic of Korea
 ZIP: 422-230
 ADDRESSEE: KIM, Chong Suh
 STREET: Garden Heights Apt. 202-801, #100,
 STREET: Hwangkeum-dong, Soosung-ku
 CITY: Taegu
 STATE: Taegu
 COUNTRY: Republic of Korea
 ZIP: 706-040
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy Disk, 3.5 inch, 1.44MB storage
 COMPUTER: IBM PC/AT
 OPERATING SYSTEM: MS-DOS
 SOFTWARE: Word Perfect 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/600,783
 FILING DATE:
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: KR 95-2751
 FILING DATE: 15-FEB-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Shahan Islam
 REGISTRATION NUMBER: 32,507
 REFERENCE/DOCKET NUMBER:
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 278-1000
 TELEFAX: (212) 953-7249
 INFORMATION FOR SEQ ID NO: 17:

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 28, 2004, 22:07:29 ; Search time 580.217 Seconds
(without alignments)
97.211 Million cell updates/sec

Title: US-10-783-095-9

Perfect score: 11

Sequence: 1 QPLALESLQK 11

Scoring table:

Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3413475 seqs, 2563800928 residues

Word size: 1

Total number of hits satisfying chosen parameters: 6814041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: listing first 45 summaries

Command line parameters:

-WDEI=frame+p2n.model -DEV=xlp
-Q=/cgn2_1/USPFO.spool.p/US10783095/runat_26102004_185456_25148/app_query.fasta_1.796
-DB=PublishedApplications NA -QMT=fastap -SUFFIX=oli.rnpp -MINMATCH=0.1
-LOOPEL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=quality -THR MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10783095 @CGN 1 1 1189 @runat_26102004_185456_25148
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOK=100
-LONGLOG -DBV_TIMEOUT=120 -WARN_TIMEOUT=30 -THRSADS=1 -XGAPOP=60 -XGAPEXT=60
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
19: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
20: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
21: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

*

Result No.	Score	Query Match	Length	DB ID	Description
1	11	100.0	258	15	US-10-383-285-1
2	11	100.0	260	10	US-09-807-742-16
3	11	100.0	260	10	US-09-807-742-17
4	11	100.0	260	17	US-10-746-149-23
5	11	100.0	260	18	US-10-609-019-21
6	11	100.0	330	15	US-10-346-563-1
7	11	100.0	330	16	US-10-321-717-1
8	11	100.0	351	9	US-09-280-030-49
9	11	100.0	350	9	US-09-280-030-48
10	11	100.0	450	9	US-09-804-409A-10
11	11	100.0	450	10	US-09-969-748C-5
12	11	100.0	450	15	US-10-328-813-1
13	11	100.0	450	16	US-10-411-037-43
14	11	100.0	450	16	US-10-411-026-43
15	11	100.0	450	16	US-10-410-962-43
16	11	100.0	450	16	US-10-411-049-43
17	11	100.0	450	17	US-10-410-930-43
18	11	100.0	450	17	US-10-410-997-43
19	11	100.0	450	17	US-10-411-012-43
20	11	100.0	450	17	US-10-287-994-43
21	11	100.0	450	17	US-10-410-913-43
22	11	100.0	1182	9	US-09-919-344-27
23	11	100.0	1626	18	US-10-609-019-27
24	11	100.0	2499	10	US-09-775-508A-1
25	11	100.0	10487	17	US-10-746-149-44
26	11	100.0	10487	18	US-10-609-019-32
27	11	100.0	10512	17	US-10-746-149-42
28	11	100.0	10512	18	US-10-609-019-31
29	11	100.0	10880	17	US-10-746-149-47
30	11	100.0	10895	18	US-10-609-019-41
31	11	100.0	11255	17	US-10-746-149-43
32	11	100.0	11271	18	US-10-609-019-42
33	11	100.0	11332	18	US-10-609-019-43
34	10	90.9	68	16	US-10-430-752A-9
35	10	90.9	74	16	US-10-430-752A-7
36	10	90.9	375	16	US-10-221-677-25
37	10	90.9	1944	16	US-10-221-677-26
38	10	90.9	4846	16	US-10-430-752A-14
39	8	72.7	333	17	US-10-419-539-2
40	8	72.7	598	13	US-10-013-032-11
41	8	72.7	633	13	US-10-013-032-12
42	8	72.7	666	13	US-10-013-032-13
43	8	72.7	1080	17	US-10-419-539-3
44	8	72.7	1217	17	US-10-419-539-4
45	7	63.6	323	9	US-09-783-590-1388

ALIGNMENTS

RESULT 1

US-10-383-285-1
; Sequence 1, Application US/10383285
; Publication No. US20030200556A1
; GENERAL INFORMATION:
; APPLICANT: STREATHFIELD, STEPHEN
; APPLICANT: HOWARD, JOHN
; TITLE OF INVENTION: PRODUCTION OF INSULIN AND INSULIN-LIKE PROTEINS IN PLANTS
; TITLE OF INVENTION: PLANTS
; FILE REFERENCE: 10040
; CURRENT APPLICATION NUMBER: US/10/383,285
; CURRENT FILING DATE: 2003-05-21
; PRIOR APPLICATION NUMBER: 60/362,874
; PRIOR FILING DATE: 2002-03-08
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 258
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-383-285-1

Alignment Scores:
Pred. No.: 0.00302 Length: 258
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 15 Gaps: 0

US-10-783-095-9 (1-11) x US-10-383-285-1 (1-258)

QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnLys 11
DB 160 CAGCGCTGCGCCTCGAGGGCTCCCTCAGAAG 192

RESULT 2

US-09-807-742-16
; Sequence 16, Application US/09807742
; Publication No. US20030204864A1
; GENERAL INFORMATION:
; APPLICANT: DANIELL, HENRY
; TITLE OF INVENTION: PRODUCTION OF PHARMACEUTICAL PROTEINS IN TRANSGENIC
; FILE OF INVENTION: PLASTIDS
; FILE REFERENCE: 1465-PCT-US-00
; CURRENT APPLICATION NUMBER: US/09/807,742
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: PCT/US01/06288
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 260
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-807-742-16

Alignment Scores:
Pred. No.: 0.00304 Length: 260
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-10-783-095-9 (1-11) x US-09-807-742-16 (1-260)

QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnLys 11
DB 160 CAGCGCTGCGCCTCGAGGGCTCCCTCAGAAG 192

RESULT 3

US-09-807-742-17
; Sequence 17, Application US/09807742
; Publication No. US20030204864A1
; GENERAL INFORMATION:
; APPLICANT: DANIELL, HENRY
; TITLE OF INVENTION: PRODUCTION OF PHARMACEUTICAL PROTEINS IN TRANSGENIC
; FILE OF INVENTION: PLASTIDS
; FILE REFERENCE: 1465-PCT-US-00
; CURRENT APPLICATION NUMBER: US/09/807,742
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: PCT/US01/06288
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 260
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Chloroplast
US-09-807-742-17

Alignment Scores:
Pred. No.: 0.00304 Length: 260
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-10-783-095-9 (1-11) x US-09-807-742-17 (1-260)

QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnLys 11
DB 160 CAACCTTAGCTTAGAAGGTTCTTTACAAAA 192

RESULT 4

US-10-746-149-23
; Sequence 23, Application US/10746149
; Publication No. US20040172667A1
; GENERAL INFORMATION:
; APPLICANT: Cooper, Richard K.
; APPLICANT: Fioretti, William C.
; APPLICANT: Cadd, Gary G.
; TITLE OF INVENTION: Administration of Transposon-Based Vectors to Reproductive Organs
; FILE REFERENCE: 51687-0280 (51687-294923)
; CURRENT APPLICATION NUMBER: US/10/746,149
; CURRENT FILING DATE: 2003-12-24
; PRIOR APPLICATION NUMBER: US 60/392,415
; PRIOR FILING DATE: 2002-06-26
; PRIOR APPLICATION NUMBER: US 60/441,381
; PRIOR FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: US 60/441,447
; PRIOR FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: US 60/441,405
; PRIOR FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: US 60/441,502
; PRIOR FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: US 60/441,392
; PRIOR FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: US 60/441,377
; PRIOR FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: US 10/609,019
; PRIOR FILING DATE: 2003-06-26
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 23
; LENGTH: 260
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-746-149-23

Alignment Scores:
Pred. No.: 0.00304 Length: 260
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 17 Gaps: 0

US-10-783-095-9 (1-11) x US-10-746-149-23 (1-260)

QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnLys 11
DB 160 CAGCCCTTGGCCCTCGAGGGTCCCTCAGAAG 192

RESULT 5

US-10-609-019-21
; Sequence 21, Application US/10609019
; Publication No. US20040197910A1
; GENERAL INFORMATION:
; APPLICANT: Cooper, Richard K.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2004, 11:13:10 ; Search time 5.6988 Seconds
(without alignments)
185.721 Million cell updates/sec

Title: US-10-783-095-9

Perfect score: 11

Sequence: 1 QPLALEGSLQK 11

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283416 segs, 96216763 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR 79:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	11	100.0	110	1	IPHU
2	11	100.0	110	2	insulin precursor
3	11	100.0	110	2	insulin precursor
4	11	100.0	110	2	insulin precursor
5	7	63.6	110	1	insulin precursor
6	7	63.6	456	2	hypothetical prote
7	6	54.5	80	2	hypothetical prote
8	6	54.5	110	1	insulin precursor
9	6	54.5	156	2	glutamine-fructose
10	6	54.5	199	2	hypothetical prote
11	6	54.5	205	2	hypothetical prote
12	6	54.5	218	2	hypothetical prote
13	6	54.5	220	2	SOJ protein (soj)
14	6	54.5	230	2	hypothetical prote
15	6	54.5	232	2	photosystem II oxy
16	6	54.5	265	2	movement protein -
17	6	54.5	267	2	probable transport
18	6	54.5	272	2	hypothetical prote
19	6	54.5	275	2	enoyl-lacyl-carrie
20	6	54.5	275	2	enoyl-acyl carrier
21	6	54.5	297	2	hypothetical prote
22	6	54.5	322	2	porphobilinogen de
23	6	54.5	344	2	O-sialoglycoprotei
24	6	54.5	344	2	O-sialoglycoprotei
25	6	54.5	358	2	probable Transposa
26	6	54.5	363	2	deoxyhypusine synt
27	6	54.5	384	2	hypothetical prote
28	6	54.5	405	2	probable FAD-depen
29	6	54.5	440	2	probable indole-3-

30	6	54.5	455	2	T00584
31	6	54.5	464	2	intermediate filam
32	6	54.5	479	2	hydrogenase-4 comp
33	6	54.5	479	2	hydrogenase 4 memb
34	6	54.5	479	2	hydrogenase 4 memb
35	6	54.5	484	2	thiazole biosynthe
36	6	54.5	495	2	NADH2 dehydrogenas
37	6	54.5	518	2	n-acetylglucosamin
38	6	54.5	524	2	ATP-dependent RNA
39	6	54.5	524	2	conserved hypotet
40	6	54.5	595	2	atp dependent heli
41	6	54.5	597	2	glutamine-fructose
42	6	54.5	597	2	glutamine-fructose
43	6	54.5	598	2	glutamine-fructose
44	6	54.5	608	2	glutamine-fructose
45	6	54.5	608	2	hypothetical prote

ALIGNMENTS

RESULT 1

insulin precursor [validated] - human
N:Alternate names: preproinsulin
C:Species: Homo sapiens (man)
C>Date: 23-Oct-1981 #sequence revision 23-Oct-1981 #text change 09-Jul-2004
C:Accession: A93222; A94253; A93216; A94251; A93144; A92075; A91186; I58114; A01579; S58
R:Bell, G.I.; Pictet, R.L.; Rutter, W.J.; Cordell, B.; Tischer, E.; Goodman, H.M.
Nature 284, 28-32, 1980
A:Title: Sequence of the human insulin gene.
A:Reference number: A93222; MUID:80120725; PMID:6243748
A:Accession: A93222
A:Molecule type: DNA
A:Residues: 1-110 <BEL>
A:Cross-references: UNIPROT:P01308; GB:J00265; NID:G186429; PIDN:AAA59172.1; PID:G386828
R:Ullrich, A.; Dull, T.J.; Gray, A.; Brosius, J.; Sures, I.
Science 209, 612-615, 1980
A:Title: Genetic variation in the human insulin gene.
A:Reference number: A94253; MUID:80236313; PMID:6248962
A:Accession: A94253
A:Molecule type: DNA
A:Residues: 1-110 <ULL>
A:Cross-references: GB:J00265; NID:G186429; PIDN:AAA59172.1; PID:G386828
R:Bell, G.I.; Swain, W.F.; Pictet, R.; Cordell, B.; Goodman, H.M.; Rutter, W.J.
Nature 282, 525-527, 1979
A:Title: Nucleotide sequence of a cDNA clone encoding human preproinsulin.
A:Reference number: A93216; MUID:80054779; PMID:503234
A:Accession: A93216
A:Molecule type: mRNA
A:Residues: 1-110 <BEL2>
A:Cross-references: GB:J00265; NID:G186429; PIDN:AAA59172.1; PID:G386828
R:Sures, I.; Goeddel, D.V.; Gray, A.; Ullrich, A.
Science 208, 57-59, 1980
A:Title: Nucleotide sequence of human preproinsulin complementary DNA.
A:Reference number: A94251; MUID:80147417; PMID:6927840
A:Accession: A94251
A:Molecule type: mRNA
A:Residues: 1-110 <SUR>
A:Cross-references: GB:J00265; NID:G186429; PIDN:AAA59172.1; PID:G386828
R:Nicol, D.S.H.W.; Smith, L.F.
Nature 187, 483-485, 1960
A:Title: Amino-acid sequence of human insulin.
A:Reference number: A93144
A:Accession: A93144
A:Molecule type: protein
A:Residues: 25-54;90-110 <NIC>
R:Over, P.E.; Cho, S.; Peterson, J.D.; Steiner, D.F.
J. Biol. Chem. 246, 1375-1386, 1971
A:Title: Studies on human proinsulin. Isolation and amino acid sequence of the human pan
A:Reference number: A92075; MUID:71116410; PMID:5101771
A:Accession: A92075
A:Molecule type: protein

```

A;Residues: 57-87 <OYE>
R;Ko, A.; Smyth, D.G.; Markussen, J.; Sundby, F.
Eur. J. Biochem. 20, 190-199, 1971
A;Title: Amino acid sequence of the C-peptide of human proinsulin.
A;Reference number: A31186; MUID:71257722; PMID:5560404
A;Accession: A91186
A;Molecule type: protein
A;Residues: 57-87 <KQ>
R;Lucassen, A.M.; Jullier, C.; Beressi, J.P.; Boitard, C.; Froguel, P.; Lathrop, M.; Bell
Nature Genet. 4, 305-310, 1993
A;Title: Susceptibility to insulin dependent diabetes mellitus maps to a 4.1 kb segment
A;Reference number: I58114; MUID:93364428; PMID:8358440
A;Accession: I58114
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-59,63-110 <RES>
A;Cross-references: GB:L15440; NID:G307071; PIDN:AAA59179.1; PID:G307072
R;Sieber, P.; Kamber, B.; Hartmann, A.; Joehli, A.; Riniker, B.; Rittel, W.
Helv. Chim. Acta 57, 2617-2621, 1974
A;Title: Totalsynthese von Humaninsulin unter gezielter Bildung der Disulfidbindungen.
A;Reference number: A31636; MUID:75077277; PMID:4443293
A;Contents: annotation; synthesis
A;Note: disulfide-bonded human insulin was synthesized; the synthetic hormone was identical
A;Note: article in German with English abstract
R;Naithani, V.K.
Hoppe-Seyler's Z. Physiol. Chem. 354, 659-672, 1973
A;Title: The synthesis of C-peptide of human proinsulin.
A;Reference number: A31658; MUID:75040007; PMID:4803504
A;Contents: annotation; synthesis of residues 57-87
R;Geiger, R.; Jaeger, G.; Koenig, W.
Chem. Ber. 106, 2347-2352, 1973
A;Title: Synthesis of the complete sequence of human proinsulin C-peptide and its [Glu-57]
A;Reference number: A30914
A;Contents: annotation; synthesis of residues 57-87
R;Kaufmann, J.E.; Irminger, J.C.; Halban, P.A.
Biochem. J. 310, 869-874, 1995
A;Title: Sequence requirements for proinsulin processing at the B-chain/C-peptide junction
A;Reference number: S58661; MUID:96013185; PMID:7575420
A;Contents: annotation; site-directed mutagenesis study of proteolytic processing
C;Genetics:
A;Gene: GDB:INS
A;Cross-references: GDB:119349; OMIM:176730
A;Map position: lip15.5-lip15.5
A;Introns: 63/1
C;Superfamily: insulin
C;Keywords: hormone; pancreas
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-54/Domain: insulin chain B #status experimental <BCH>
F;25-54,90-110/Product: insulin #status predicted <MAT>
F;57-87/Domain: connecting C peptide #status experimental <CPEP>
F;90-110/Domain: insulin chain A #status experimental <ACH>
F;31-96,43-109,95-100/Disulfide bonds: #status experimental
Query Match 100.0%; Score 11; DB 1; Length 110;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPLALEGSLQK 11
DB 78 QPLALEGSLQK 88

RESULT 2
insulin precursor - green monkey
C;Species: Cercopithecus aethiops (green monkey, grivet)
C;Date: 04-Mar-1993 #sequence revision 18-Nov-1994 #text_change 09-Jul-2004
C;Accession: B42179; A05232; S16494; S22056
R;Seino, S.; Bell, G.I.; Li, W.H.
Mol. Biol. Evol. 9, 193-203, 1992
A;Title: Sequences of primate insulin genes support the hypothesis of a slower rate of
A;Reference number: A42179; MUID:92219953; PMID:1560757
A;Accession: B42179

RESULT 3
insulin precursor - crab-eating macaque
C;Species: Macaca fascicularis (crab-eating macaque)
C;Date: 07-Sep-1990 #sequence revision 07-Sep-1990 #text_change 09-Jul-2004
C;Accession: JQ0178
R;Retekam, W.; Gronenberg, J.; Leineweber, M.; Wengenmayer, F.; Winnacker, E.L.
Gene 19, 179-183, 1982
A;Title: The nucleotide sequence of cDNA coding for proinsulin from the primate Macaca
A;Reference number: JQ0178; MUID:83080474; PMID:6184262
A;Accession: JQ0178
A;Molecule type: mRNA
A;Residues: 1-110 <WET>
A;Cross-references: UNIPROT:P30406; GB:J00336; NID:G342121; PIDN:AAA36849.1; PID:G342122
C;Superfamily: insulin
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-54,90-110/Product: insulin #status predicted <MAT>
F;25-54/Domain: insulin chain B #status predicted <BCH>
F;55-89/Domain: insulin connecting C peptide #status predicted <CPT>
F;90-110/Domain: insulin chain A #status predicted <ACH>
F;31-96,43-109,95-100/Disulfide bonds: #status predicted
Query Match 100.0%; Score 11; DB 2; Length 110;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPLALEGSLQK 11
DB 78 QPLALEGSLQK 88

RESULT 4
insulin precursor - chimpanzee
C;Species: Pan troglodytes (chimpanzee)
C;Date: 04-Mar-1993 #sequence revision 18-Nov-1994 #text_change 09-Jul-2004
C;Accession: A42179; S22058
R;Seino, S.; Bell, G.I.; Li, W.H.
Mol. Biol. Evol. 9, 193-203, 1992
A;Title: Sequences of primate insulin genes support the hypothesis of a slower rate of
A;Reference number: A42179; MUID:92219953; PMID:1560757
A;Accession: A42179

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 28, 2004, 19:58:09 ; Search time 1249.89 Seconds
(without alignments)
320.697 Million cell updates/sec

Title: US-10-783-095-9
Perfect score: 11
Sequence: 1 QPLAEGSLQK 11

Scoring table: OLIGO
Xgapop 60.0, Xgapext 60.0
Ygapop 60.0, Ygapext 60.0
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 32822875 seqs, 18219865908 residues
Word size: 1

Total number of hits satisfying chosen parameters: 65641631

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame p2n.model -DEV=xlp
-O=/cpn2.1/USPTO.spool.p/US10783095/runat_26102004.185454.25046/app_query.fasta_1.796
-DB=EST_QPWT=fastap SUPFIX=oli.rst -MINMATCH=0.1 -LOOPCL=0 -LOPEXT=45
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFWT=ptc
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10783095 @GCN 1 1 9525 @runat_26102004.185454.25046 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : EST:
1: gb_est1:
2: gb_est2:
3: gb_hic:
4: gb_est3:
5: gb_est4:
6: gb_est5:
7: gb_est6:
8: gb_gss1:
9: gb_gss2:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	11	100.0	118	5	BQ267254
C 2	11	100.0	118	5	BQ272508
C 3	11	100.0	118	5	BU073408
C 4	11	100.0	150	4	BI466839
C 5	11	100.0	150	4	BI466930
C 6	11	100.0	162	5	BU074185
C 7	11	100.0	166	4	BI711381
C 8	11	100.0	168	4	BM511718
C 9	11	100.0	173	5	BU073285

C 10	11	100.0	174	5	BQ272395
C 11	11	100.0	176	6	CA948611
C 12	11	100.0	178	4	BM510932
C 13	11	100.0	179	5	BQ548672
C 14	11	100.0	185	4	BM857672
C 15	11	100.0	189	5	BQ267409
C 16	11	100.0	190	4	BM857847
C 17	11	100.0	190	5	BU787471
C 18	11	100.0	191	5	BQ270646
C 19	11	100.0	191	5	BU071032
C 20	11	100.0	191	5	BU075717
C 21	11	100.0	191	5	BU077451
C 22	11	100.0	194	4	BM565510
C 23	11	100.0	197	4	BI711514
C 24	11	100.0	197	5	BQ270135
C 25	11	100.0	198	4	BM565445
C 26	11	100.0	198	5	BU784487
C 27	11	100.0	198	5	BU790178
C 28	11	100.0	199	4	BI711217
C 29	11	100.0	199	5	BU069976
C 30	11	100.0	199	5	BU071888
C 31	11	100.0	199	5	BU789127
C 32	11	100.0	200	5	BU579459
C 33	11	100.0	203	4	BI711616
C 34	11	100.0	203	4	BM506695
C 35	11	100.0	203	4	BM507174
C 36	11	100.0	203	4	BM511187
C 37	11	100.0	203	5	BU070221
C 38	11	100.0	203	5	BU789388
C 39	11	100.0	204	5	BU071687
C 40	11	100.0	208	5	BU786529
C 41	11	100.0	216	4	BM509111
C 42	11	100.0	220	5	BQ286797
C 43	11	100.0	220	5	BU784792
C 44	11	100.0	220	6	CA948576
C 45	11	100.0	221	5	BQ271297

ALIGNMENTS

RESULT 1
LOCUS BQ267254/c
DEFINITION BQ267254.1 Human insulinoma Homo sapiens cDNA clone IMAGE:5779548
5' similar to SW:INS_HUMAN F01308 INSULIN PRECURSOR. [1] ; MRNA
sequence.
BQ267254.1 GI:20492319
EST.
Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 118)
Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
Hillier,L., Marra,M., Page,D., Wylie,T., Martin,J., Blistain,A.,
Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J.,
Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagarisshvili,R.,
Williams,T., Jackson,Y. and Bowers,Y.
Endocrine Pancreas Consortium
Unpublished (2000)
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biochem.harvard.edu
Library was constructed by Dr. J. Ferrer In vivo mass-excised to
pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington
University Genome Sequencing Center For information on obtaining a

BQ267254 118 bp MRNA linear EST 15-JUL-2003
ik01b07.y1 Human insulinoma Homo sapiens cDNA clone IMAGE:5779548
5' similar to SW:INS_HUMAN F01308 INSULIN PRECURSOR. [1] ; MRNA
sequence.
BQ267254.1 GI:20492319
EST.
Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 118)
Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
Hillier,L., Marra,M., Page,D., Wylie,T., Martin,J., Blistain,A.,
Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J.,
Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagarisshvili,R.,
Williams,T., Jackson,Y. and Bowers,Y.
Endocrine Pancreas Consortium
Unpublished (2000)
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biochem.harvard.edu
Library was constructed by Dr. J. Ferrer In vivo mass-excised to
pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington
University Genome Sequencing Center For information on obtaining a

clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
Possible reversed clone: similarity on wrong strand
Seq primer: -40RP from Gibco.

FEATURES

source

```
1..118
  /organism="Homo sapiens"
  /mol_type="mRNA"
  /db_xref="taxon:9606"
  /clone="IMAGE:5779548"
  /tissue_type="insulinoma"
  /lab_host="DH10B (phage-resistant)"
  /clone_lib="Human insulinoma"
  /note="Organ: pancreas; Vector: pBluescript SK-; Site 1:
XhoI; Site 2: EcoRI; Constructed with lambda ZAPII system
(Stratagene) by Dr. J. Ferrer, in vivo mass-excised to
pBluescript SK- by Dr. H. Inoue following the Washington
University protocol
(http://genome.wustl.edu/est/lambda_protocol.shtml).
Please contact Hiroshi Inoue, MD/PhD for further
information on this library (Metabolism Division, Permutt
Laboratory, Washington University School of Medicine, Box
8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this
is a Washington University Pancreas EST project library."
```

ORIGIN

Alignment Scores:
Pred. No.: 0.0897 Length: 118
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 5 Gaps: 0

US-10-783-095-9 (1-11) x BQ267254 (1-118)

```
QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnLys 11
DB 113 CAGCCCTTGGCCCTGGAGGGTCCCTGCAGAAG 81
```

RESULT 2
LOCUS BQ272508 118 bp mRNA linear EST 15-JUL-2003
DEFINITION ik01b07.x1 Human insulinoma Homo sapiens cDNA clone IMAGE:5779548
3' similar to SWINS_HUMAN P01308 INSULIN PRECURSOR. [1] ; mRNA
sequence.

ACCESSION BQ272508
VERSION BQ272508.1 GI:20497577
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J.,
Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R.,
Williams,T., Jackson,Y. and Bowers,Y.

Unpublished Pancreas Consortium
Other ESTs: ik01b07.y1
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu

Library was constructed by Dr. J. Ferrer in vivo mass-excised to
pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington
University Genome Sequencing Center For information on obtaining a
clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)

clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
Possible reversed clone: similarity on wrong strand
Seq primer: -40UP from Gibco.

FEATURES

source

```
1..118
  /organism="Homo sapiens"
  /mol_type="mRNA"
  /db_xref="taxon:9606"
  /clone="IMAGE:5779548"
  /tissue_type="insulinoma"
  /lab_host="DH10B (phage-resistant)"
  /clone_lib="Human insulinoma"
  /note="Organ: pancreas; Vector: pBluescript SK-; Site 1:
XhoI; Site 2: EcoRI; Constructed with lambda ZAPII system
(Stratagene) by Dr. J. Ferrer, in vivo mass-excised to
pBluescript SK- by Dr. H. Inoue following the Washington
University protocol
(http://genome.wustl.edu/est/lambda_protocol.shtml).
Please contact Hiroshi Inoue, MD/PhD for further
information on this library (Metabolism Division, Permutt
Laboratory, Washington University School of Medicine, Box
8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this
is a Washington University Pancreas EST project library."
```

ORIGIN

Alignment Scores:
Pred. No.: 0.0897 Length: 118
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 5 Gaps: 0

US-10-783-095-9 (1-11) x BQ272508 (1-118)

```
QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnLys 11
DB 5 CAGCCCTTGGCCCTGGAGGGTCCCTGCAGAAG 37
```

RESULT 3
LOCUS BU073408 118 bp mRNA linear EST 27-AUG-2002
DEFINITION im35e06.x1 Human insulinoma Homo sapiens cDNA clone IMAGE:6036971
3' similar to SWINS_HUMAN P01308 INSULIN PRECURSOR. [1] ; mRNA
sequence.

ACCESSION BU073408
VERSION BU073408.1 GI:22514597
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J.,
Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R.,
Williams,T., Jackson,Y. and Bowers,Y.

Unpublished Pancreas Consortium
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu

Library was constructed by Dr. J. Ferrer in vivo mass-excised to
pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington
University Genome Sequencing Center For information on obtaining a
clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)

OM protein - protein search, using sw model

Run on: October 28, 2004, 10:57:30 ; Search time 27.8313 Seconds
(without alignments) 227.410 Million cell updates/sec

Title: US-10-783-095-9
perfect score: 11
Sequence: 1 OPLALEGSLOK 11

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1825181 seqs, 575374646 residues

Word size : 0

Total number of hits satisfying chosen parameters:	1825181
--	---------

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Listing first 45 summaries

```
Database :      UniProt_02:*
            1: uniprot_sprot:*
            2: uniprot_trembl:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query %			DB	ID	Description
	Score	Match	Length			
1	11	100.0	65	2	Q8HZ80	Q8hz80 pongo pygma
2	11	100.0	65	2	Q8HZ81	Q8hz81 gorilla gor
3	11	100.0	110	1	INS_CSERAE	P30407 cercopithec
4	11	100.0	110	1	INS_HUMAN	P01308 homo sapien
5	11	100.0	110	1	INS_MACFA	P30406 macaca fasc
6	11	100.0	110	1	INS_PANTR	P30410 pan troglod
7	11	100.0	110	1	INS_PONPY	Q8hxv2 pongo pygma
8	11	100.0	110	2	Q6YK33	Q6yK33 gorilla gor
9	11	100.0	110	2	AAP35454	AAP35454 homo sapi
10	11	100.0	110	2	AAM06935	AAM06935 gorilla g
11	11	100.0	110	2	AAM39451	AAM39451 homo sapi
12	7	63.6	110	1	INS_CANFA	P01321 canis famli
13	7	63.6	188	2	Q88AB3	Q88ab3 pseudomoras
14	7	63.6	456	2	O59732	O59732 schizosacch
15	7	63.6	497	1	ATRX_MACEU	P83798 macropus eu
16	7	63.6	837	2	Q7MTD5	Q7mtd5 porphyromon
17	7	63.6	1395	1	CGB3_HUMAN	Q8aw17 homo sapien
18	7	63.6	2592	2	Q84TL0	Q84tl0 oryza sativ
19	6	54.5	62	1	MLC2_HYLLA	O62694 hylobates l
20	6	54.5	62	1	MLC2_PANPA	O62692 pan paniscu
21	6	54.5	62	1	MLC2_PANTR	O62693 pan troglod
22	6	54.5	72	2	Q9NRF0	Q9nrf0 homo sapien
23	6	54.5	78	1	MLC2_HUMAN	Q16048 homo sapien
24	6	54.5	80	2	Q8Y957	Q8y9b7 listeria mo
25	6	54.5	80	2	Q722T5	Q722t5 listeria mo
26	6	54.5	80	2	AAT03426	AAT03426 listeria
27	6	54.5	82	2	Q9BQC4	Q9bqc4 homo sapien
28	6	54.5	83	2	Q94ZY3	Q94zy3 oryza sativ
29	6	54.5	86	2	Q9BQ10	Q9bq10 homo sapien
30	6	54.5	86	2	Q9BQD1	Q9bqd1 homo sapien
31	6	54.5	86	2	Q89JU2	Q89ju2 bradyrhizob

DR GO; GO:0005576; C:extracellular; IEA.
 DR GO; GO:0005179; F:hormone activity; IEA.
 DR GO; GO:0007582; P:physiological process; IEA.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR Pfam; PF00049; Insulin; 1.
 DR SMART; SM00078; IIGF; 1.
 KW Insulin family.
 FT NON TER 1
 FT NON TER 65
 SQ SEQUENCE 65 AA; 6920 MW; B772017FD8BCABEA CRC64;
 Query Match 100.0%; Score 11; DB 2; Length 65;
 Best Local Similarity 100.0%; Pred. No. 0.001;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 QPLALEGSLQK 11
 Db 48 QPLALEGSLQK 58

RESULT 3
 INS_CERAE
 ID INS_CERAE STANDARD; PRT; 110 AA.
 AC P0307; P01309;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Insulin precursor.
 GN Name=INS;
 OS Cercopithecus aethiops (Green monkey) (Grivet).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecinae; Cercopithecus.
 OX NCBI_TaxID=9534;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92219953; PubMed=1560757;
 RA Seino S., Bell G.I., Li W.;
 RT "Sequences of primate insulin genes support the hypothesis of a slower
 rate of molecular evolution in humans and apes than in monkeys.";
 RL Mol. Biol. Evol. 9:193-203(1992).
 RN [2]
 RP SEQUENCE OF 57-87.
 RX MEDLINE=72258016; PubMed=4626369;
 RA Peterson J.D., Nehrlich S., Oyer P.E., Steiner D.P.;
 RT "Determination of the amino acid sequence of the monkey, sheep, and
 dog proinsulin C-peptides by a semi-micro Edman degradation
 procedure.";
 RL J. Biol. Chem. 247:4866-4871(1972).
 CC -!- FUNCTION: Insulin decreases blood glucose concentration. It
 increases cell permeability to monosaccharides, amino acids and
 fatty acids. It accelerates glycolysis, the pentose phosphate
 cycle, and glycogen synthesis in liver.
 CC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
 disulfide bonds.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: Belongs to the insulin family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X61092; CAA43405.1; --
 DR PIR; B42179; B42179.
 DR HSSP; P01308; 1A10.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR Pfam; PF00049; Insulin; 1.
 DR PRINTS; PR00277; INSULIN.
 DR SMART; SM00078; IIGF; 1.

DR PROSITE; PS00262; INSULIN; 1.
 KW Direct protein sequencing; Glucose metabolism; Hormone;
 FT SIGNAL 1 24
 FT CHAIN 25 54 Insulin B chain.
 FT PROPEP 57 87 C peptide.
 FT CHAIN 90 110 Insulin A chain.
 FT DISULFID 31 96 Interchain.
 FT DISULFID 43 109 Interchain.
 FT DISULFID 95 100
 SQ SEQUENCE 110 AA; 12019 MW; 95A1F54BE7B247F9 CRC64;
 Query Match 100.0%; Score 11; DB 1; Length 110;
 Best Local Similarity 100.0%; Pred. No. 0.0016;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 QPLALEGSLQK 11
 Db 78 QPLALEGSLQK 88

RESULT 4
 INS_HUMAN
 ID INS_HUMAN STANDARD; PRT; 110 AA.
 AC P01308;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 01-OCT-2004 (Rel. 45, Last annotation update)
 DE Insulin precursor.
 GN Name=INS;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=80120725; PubMed=6243748;
 RA Bell G.I., Pictet R.L., Rutter W.J., Cordell B., Tischner E.,
 RA Goodman H.M.;
 RT "Sequence of the human insulin gene.";
 RL Nature 284:26-32(1980).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=80236313; PubMed=6248962;
 RA Ullrich A., Dull T.J., Gray A., Brosius J., Sures I.;
 RT "Genetic variation in the human insulin gene.";
 RL Science 209:612-615(1980).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=80054779; PubMed=503234;
 RA Bell G.I., Swain W.F., Pictet R.L., Cordell B., Goodman H.M.,
 RA Rutter W.J.;
 RT "Nucleotide sequence of a cDNA clone encoding human preproinsulin.";
 RL Nature 282:525-527(1979).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=80147417; PubMed=6927840;
 RA Sures I., Goeddel D.V., Gray A., Ullrich A.;
 RT "Nucleotide sequence of human preproinsulin complementary DNA.";
 RL Science 208:57-59(1980).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93364428; PubMed=8358440;
 RA Lucassen A.M., Bell J.I., Jullier C., Lathrop M.;
 RT "Susceptibility to insulin dependent diabetes mellitus maps to a 4.1
 kb segment of DNA spanning the insulin gene and associated VNTR.";
 RL Nat. Genet. 4:305-310(1993).
 RN [6]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Pancreas;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2004, 10:22:24 ; Search time 11.6627 Seconds
(without alignments)
62.550 Million cell updates/sec

Title: US-10-783-095-9
Perfect score: 53
Sequence: 1 QPLALEGSLQK 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA*
1: /cgn2_6/ptodata/1/aaa/5A-COMB.pap:*
2: /cgn2_6/ptodata/1/aaa/5B-COMB.pap:*
3: /cgn2_6/ptodata/1/aaa/6A-COMB.pap:*
4: /cgn2_6/ptodata/1/aaa/6B-COMB.pap:*
5: /cgn2_6/ptodata/1/aaa/PCTUS-COMB.pap:*
6: /cgn2_6/ptodata/1/aaa/backfiles1.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	53	100.0	23	3	US-08-105-904B-10
2	53	100.0	23	3	US-08-114-877A-10
3	53	100.0	23	3	US-08-537-928-10
4	53	100.0	23	4	US-09-633-269-10
5	53	100.0	27	2	US-08-508-664-17
6	53	100.0	29	4	US-09-676-787-2
7	53	100.0	32	6	5514646-5
8	53	100.0	33	4	US-09-639-667-24
9	53	100.0	33	6	5514646-4
10	53	100.0	35	1	US-08-160-376A-3
11	53	100.0	35	1	US-08-389-487-6
12	53	100.0	35	2	US-09-134-836-3
13	53	100.0	35	3	US-09-386-303A-3
14	53	100.0	35	4	US-09-815-229-4
15	53	100.0	35	4	US-09-947-563-3
16	53	100.0	35	6	5457066-1
17	53	100.0	54	4	US-09-815-229-13
18	53	100.0	86	1	US-07-715-183C-1
19	53	100.0	86	4	US-09-477-924-2
20	53	100.0	86	4	US-09-723-981-2
21	53	100.0	86	4	US-09-723-896-2
22	53	100.0	86	4	US-09-878-380-1
23	53	100.0	90	1	US-08-030-731A-43
24	53	100.0	91	4	US-09-676-787-7
25	53	100.0	96	2	US-09-134-836-4
26	53	100.0	96	2	US-09-134-836-5
27	53	100.0	96	3	US-09-386-303A-4

28	53	100.0	96	3	US-09-386-303A-5	Sequence 5, Appli
29	53	100.0	96	4	US-09-947-563-4	Sequence 4, Appli
30	53	100.0	96	4	US-09-947-563-5	Sequence 5, Appli
31	53	100.0	97	1	US-08-160-376A-4	Sequence 4, Appli
32	53	100.0	97	1	US-08-389-487-7	Sequence 7, Appli
33	53	100.0	97	3	US-09-099-307-6	Sequence 6, Appli
34	53	100.0	97	3	US-09-099-307-7	Sequence 7, Appli
35	53	100.0	97	3	US-09-099-307-11	Sequence 8, Appli
36	53	100.0	97	3	US-09-099-307-8	Sequence 11, Appli
37	53	100.0	98	4	US-09-701-968-7	Sequence 7, Appli
38	53	100.0	99	4	US-09-701-968-8	Sequence 8, Appli
39	53	100.0	100	4	US-09-701-968-9	Sequence 9, Appli
40	53	100.0	110	3	US-08-950-720A-11	Sequence 11, Appli
41	53	100.0	110	3	US-08-589-028-2	Sequence 2, Appli
42	53	100.0	110	3	US-08-589-028-4	Sequence 4, Appli
43	53	100.0	110	3	US-08-784-582-2	Sequence 2, Appli
44	53	100.0	110	3	US-08-784-582-4	Sequence 4, Appli
45	53	100.0	110	3	US-08-785-271-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-105-904B-10
; Sequence 10, Application US/08105904B
; Patent No. 6001364
; GENERAL INFORMATION:
; APPLICANT: Rose, Keith
; APPLICANT: Offord, Robin
; TITLE OF INVENTION: HETERO-POLYOXIME COMPOUNDS AND THEIR
; PREPARATION BY PARALLEL ASSEMBLY
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooley Godward Castro Huddleson & Tatum
; STREET: 5 Palo Alto Square, 3000 El Camino Real
; CITY: Palo Alto
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/105,904B
; FILING DATE: 31-AUG-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/057,594
; FILING DATE: 05-MAY-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Neeley, Richard L.
; REGISTRATION NUMBER: 30,092
; REFERENCE/DOCKET NUMBER: ABIC-001/02US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)843-5000
; TELEFAX: (415)857-0663
; TELEX: 380816 CooleyPA
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 23 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 1
; OTHER INFORMATION: AOA-Glu
; US-08-105-904B-10

Query Match 100.0%; Score 53; DB 3; Length 23;
 Best Local Similarity 100.0%; Pred. No. 0.00037;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPLALEGSLQK 11
 |||||
 Db 12 QPLALEGSLQK 22

RESULT 2
 US-08-114-877A-10
 ; Sequence 10, Application US/08114877A
 ; Patent No. 6174530
 ; GENERAL INFORMATION:
 ; APPLICANT: Rose, Keith
 ; APPLICANT: Offord, Robin
 ; TITLE OF INVENTION: HOMOGENOUS POLYOXIME COMPOSITIONS AND THEIR
 ; TITLE OF INVENTION: PREPARATION BY PARALLEL ASSEMBLY
 ; NUMBER OF SEQUENCES: 15
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Cooley Godward Castro Huddleson & Tatum
 ; STREET: 5 Palo Alto Square
 ; CITY: Palo Alto
 ; STATE: California
 ; COUNTRY: U.S.A.
 ; ZIP: 94036

COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/114,877A
 ; FILING DATE: 31-AUG-1993
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/057,594
 ; FILING DATE: 05-MAY-1993
 ; CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:
 ; NAME: Neeley, Richard L.
 ; REGISTRATION NUMBER: 30,092
 ; REFERENCE/DOCKET NUMBER: ASIC-001/01US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 843 5070
 ; TELEFAX: (415) 857-0663
 ; TELEX: 380816 CooleyPA
 ; INFORMATION FOR SEQ ID NO: 10:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 23 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; HYPOTHEetical: NO
 ; FEATURE:
 ; NAME/KEY: Modified-site
 ; LOCATION: 1
 ; OTHER INFORMATION: AOA-Glu
 ; US-08-114-877A-10

Query Match 100.0%; Score 53; DB 3; Length 23;
 Best Local Similarity 100.0%; Pred. No. 0.00037;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPLALEGSLQK 11
 |||||
 Db 12 QPLALEGSLQK 22

RESULT 3
 US-08-537-928-10
 ; Sequence 10, Application US/08537928

Patent No. 6217873
 ; GENERAL INFORMATION:
 ; APPLICANT: Keith, Rose
 ; APPLICANT: Offord, Robin E
 ; TITLE OF INVENTION: POLYOXIME COMPOSITIONS AND THEIR PREPARATION
 ; FILE REFERENCE: GRFN-001/03US
 ; CURRENT APPLICATION NUMBER: US/08/537,928
 ; CURRENT FILING DATE: 1996-01-05
 ; EARLIER APPLICATION NUMBER: PCT/IB94/00093
 ; EARLIER FILING DATE: 1994-05-05
 ; EARLIER APPLICATION NUMBER: 08/105,904
 ; EARLIER FILING DATE: 1993-08-31
 ; EARLIER APPLICATION NUMBER: 08/114,877
 ; EARLIER FILING DATE: 1993-08-31
 ; NUMBER OF SEQ ID NOS: 30
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 10
 ; LENGTH: 23
 ; TYPE: PRT

ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 ; NAME/KEY: SITE
 ; LOCATION: (1)
 ; OTHER INFORMATION: Aminoxyacetyl-Glu
 ; US-08-537-928-10

Query Match 100.0%; Score 53; DB 3; Length 23;
 Best Local Similarity 100.0%; Pred. No. 0.00037;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPLALEGSLQK 11
 |||||
 Db 12 QPLALEGSLQK 22

RESULT 4
 US-09-633-269-10
 ; Sequence 10, Application US/09633269
 ; Patent No. 6663869
 ; GENERAL INFORMATION:
 ; APPLICANT: Keith, Rose
 ; APPLICANT: Offord, Robin E
 ; TITLE OF INVENTION: POLYOXIME COMPOSITIONS AND THEIR PREPARATION
 ; FILE REFERENCE: GRFN-001/03US
 ; CURRENT APPLICATION NUMBER: US/09/633,269
 ; CURRENT FILING DATE: 2000-08-04
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/08/537,928
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1996-01-05
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PCT/IB94/00093
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1994-05-05
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 08/105,904
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1993-08-31
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 08/114,877
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1993-08-31
 ; NUMBER OF SEQ ID NOS: 30
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 10
 ; LENGTH: 23
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 ; NAME/KEY: SITE
 ; LOCATION: (1)
 ; OTHER INFORMATION: Aminoxyacetyl-Glu
 ; US-09-633-269-10

Query Match 100.0%; Score 53; DB 4; Length 23;
 Best Local Similarity 100.0%; Pred. No. 0.00037;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2004, 10:54:39 ; Search time 143.795 Seconds
(without alignments)
24.802 Million cell updates/sec

Title: US-10-783-095-9
Perfect score: 53
Sequence: 1 QPLALEGSLOK 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1370721 seqs, 324215800 residues

Total number of hits satisfying chosen parameters: 1370721

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53	100.0	18	17	US-10-617-568-23
2	53	100.0	33	17	US-10-641-834-24
3	53	100.0	35	9	US-09-815-229-4
4	53	100.0	35	9	US-09-947-563-3
5	53	100.0	35	16	US-10-740-098-4
6	53	100.0	54	9	US-09-815-229-13
7	53	100.0	54	16	US-10-740-098-13
8	53	100.0	86	9	US-09-878-380-1
9	53	100.0	86	10	US-09-858-935B-4
10	53	100.0	86	13	US-10-028-410-2
11	53	100.0	86	13	US-10-054-873-4
12	53	100.0	86	14	US-10-444-326-2
13	53	100.0	86	14	US-10-271-869-4

14	53	100.0	86	15	US-10-444-262-2	Sequence 2, Appli
15	53	100.0	86	15	US-10-444-649-2	Sequence 2, Appli
16	53	100.0	86	15	US-10-444-701-2	Sequence 2, Appli
17	53	100.0	96	9	US-09-947-563-4	Sequence 4, Appli
18	53	100.0	96	9	US-09-947-563-5	Sequence 5, Appli
19	53	100.0	110	9	US-09-205-658-125	Sequence 125, App
20	53	100.0	110	9	US-09-815-229-3	Sequence 3, Appli
21	53	100.0	110	9	US-09-804-409A-9	Sequence 9, Appli
22	53	100.0	110	10	US-09-969-748C-6	Sequence 6, Appli
23	53	100.0	110	10	US-09-963-693-125	Sequence 125, App
24	53	100.0	110	14	US-10-038-686-1	Sequence 1, Appli
25	53	100.0	110	14	US-10-328-813-2	Sequence 2, Appli
26	53	100.0	110	14	US-10-383-285-2	Sequence 2, Appli
27	53	100.0	110	14	US-10-346-563-2	Sequence 2, Appli
28	53	100.0	110	15	US-10-321-717-2	Sequence 2, Appli
29	53	100.0	110	15	US-10-411-037-44	Sequence 44, Appli
30	53	100.0	110	15	US-10-411-026-44	Sequence 44, Appli
31	53	100.0	110	15	US-10-410-962-44	Sequence 44, Appli
32	53	100.0	110	15	US-10-411-049-44	Sequence 44, Appli
33	53	100.0	110	15	US-10-700-725-20	Sequence 20, Appli
34	53	100.0	110	16	US-10-410-930-44	Sequence 44, Appli
35	53	100.0	110	16	US-10-410-997-44	Sequence 44, Appli
36	53	100.0	110	16	US-10-411-012-44	Sequence 44, Appli
37	53	100.0	110	16	US-10-287-994-44	Sequence 44, Appli
38	53	100.0	110	16	US-10-740-098-3	Sequence 3, Appli
39	53	100.0	110	16	US-10-410-913-44	Sequence 44, Appli
40	53	100.0	117	9	US-09-280-030-63	Sequence 63, Appli
41	53	100.0	130	9	US-09-280-030-62	Sequence 62, Appli
42	50	94.3	521	15	US-10-430-752A-5	Sequence 5, Appli
43	48	90.6	31	9	US-09-269-439-1	Sequence 1, Appli
44	48	90.6	31	9	US-09-878-380-2	Sequence 2, Appli
45	48	90.6	31	15	US-10-430-752A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-10-617-568-23
; Sequence 23, Application US/10617568
; Publication No. US20040197862A1
; GENERAL INFORMATION:
; APPLICANT: Dana-Farber Cancer Institute, Inc. et al.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: GENERATION OF MHC CLASS II COMPOUNDS BY
; TITLE OF INVENTION: PEPTIDE EXCHANGE
; FILE REFERENCE: DFN-044
; CURRENT APPLICATION NUMBER: US/10/617,568
; CURRENT FILING DATE: 2003-07-11
; PRIOR APPLICATION NUMBER: 60/395494
; PRIOR FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 60/397893
; PRIOR FILING DATE: 2002-07-22
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-617-568-23

Query Match 100.0%; Score 53; DB 17; Length 18;
Best Local Similarity 100.0%; Pred. No. 0.0021;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QPLALEGSLOK 11
Db 6 QPLALEGSLOK 16

RESULT 2
US-10-641-834-24
; Sequence 24, Application US/10641834

```

; Publication No. US20040209830A1
; GENERAL INFORMATION: Stephen James
; APPLICANT: Russell, Stephen
; TITLE OF INVENTION: SYSTEM FOR MONITORING THE EXPRESSION OF
; TRANSGENES
; FILE REFERENCE: 07039-292002
; CURRENT APPLICATION NUMBER: US/10/641,834
; PRIOR FILING DATE: 2003-05-18
; PRIOR APPLICATION NUMBER: 09/639,667
; PRIOR FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: 60/149,168
; PRIOR FILING DATE: 1999-08-17
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 33
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: C-peptide sequence
US-10-641-834-24

Query Match 100.0%; Score 53; DB 17; Length 33;
Best Local Similarity 100.0%; Pred. No. 0.0041;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPLALEGSLQK 11
DB 22 QPLALEGSLQK 32

RESULT 3
US-09-815-229-4
; Sequence 4, Application US/09815229
; Patent No. US20020058614A1
; GENERAL INFORMATION:
; APPLICANT: Filvaroff, Ellen H.
; APPLICANT: Okumu, Franklin W.
; TITLE OF INVENTION: USE OF INSULIN FOR THE TREATMENT OF CARTILAGINOUS DISORDERS
; FILE REFERENCE: P178681US
; CURRENT APPLICATION NUMBER: US/09/815,229
; CURRENT FILING DATE: 2001-03-22
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/192,103
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 4
; LENGTH: 35
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-815-229-4

Query Match 100.0%; Score 53; DB 9; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.0044;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPLALEGSLQK 11
DB 24 QPLALEGSLQK 34

RESULT 4
US-09-947-563-3
; Sequence 3, Application US/09947563
; Patent No. US20020156234A1
; GENERAL INFORMATION:
; APPLICANT: Rubroder, Franz-Josef
; APPLICANT: Keller, Reinhold
; TITLE OF INVENTION: Improved process for obtaining
; insulin precursors having correctly bonded cystine bridges
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farrah, Garrett &

```

```

; Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/947,563
; FILING DATE: 07-Sep-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/134,836
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Leslie McDonnell
; REGISTRATION NUMBER: 34,872
; REFERENCE/DOCKET NUMBER: 02481.1600-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..35
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-947-563-3

Query Match 100.0%; Score 53; DB 9; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.0044;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPLALEGSLQK 11
DB 24 QPLALEGSLQK 34

RESULT 5
US-10-740-098-4
; Sequence 4, Application US/10740098
; Publication No. US20040138101A1
; GENERAL INFORMATION:
; APPLICANT: Filvaroff, Ellen H.
; APPLICANT: Okumu, Franklin W.
; TITLE OF INVENTION: USE OF INSULIN FOR THE TREATMENT OF CARTILAGINOUS DISORDERS
; FILE REFERENCE: P178681US
; CURRENT APPLICATION NUMBER: US/10/740,098
; CURRENT FILING DATE: 2003-12-17
; PRIOR FILING DATE: 2003-12-17
; PRIOR APPLICATION NUMBER: US/09/815,229
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/192,103
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 4
; LENGTH: 35
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-740-098-4

Query Match 100.0%; Score 53; DB 16; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.0044;

```

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	53	100.0	68	6	AR077834	Sequence
2	53	100.0	87	6	AX107018	Sequence
3	53	100.0	105	12	XXHSIN	V01461 Synthetic
4	53	100.0	216	6	AX179492	Sequence

US-10-783-095-9 (1-11) x AR077834 (1-68)

QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnLys 11
DB 7 CAGCCGTTGGCGCTGGAAGGTTCCCTGCAGAA 39

RESULT 2
AX107018
LOCUS AX107018 87 bp DNA linear PAT 30-APR-2001
DEFINITION Sequence 1 from Patent WO0125278.
ACCESSION AX107018
VERSION AX107018.1 GI:13922551
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1
AUTHORS Habermann,P., Ertl,J., Meiwes,J. and Seipke,G.
TITLE C-peptide for the improved production of insulin and insulin analogues
JOURNAL Patent: WO 0125278-A 1 12-APR-2001;
FEATURES Location/Qualifiers
source 1..87
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="DNA kodierend fur C-PEPTIDEVARIANTE"

ORIGIN

Alignment Scores:
Pred. No.: 0.00821 Length: 87
Score: 53.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-783-095-9 (1-11) x AX107018 (1-87)

QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnLys 11
DB 52 CAGCCCTTGGCGCTGGAAGGTTCCCTGCAGAG 84

RESULT 3
XXHSIN
LOCUS XXHSIN 105 bp DNA linear SYN 04-AUG-1995
DEFINITION Synthetic gene for human insulin C-chain.
ACCESSION V01461 J02546
VERSION V01461.1 GI:58303
KEYWORDS artificial gene; gene synthesis.
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1 (bases 1 to 105)
AUTHORS Hsiung,H.M., Sung,W.L., Brousseau,R., Wu,R. and Narang,S.A.
TITLE Synthesis of the human insulin gene, Part III. Chemical synthesis of 5'-phosphonomoester group containing deoxyribonucleotides by the modified phosphotriester method. Its application in the synthesis of seventeen fragments constituting human insulin C-chain DNA

Nucleic Acids Res. 8 (23), 5753-5765 (1980)

JOURNAL
MEDLINE 81124297
PUBMED 7008029

FEATURES
source 1..105
Location/Qualifiers
/organism="synthetic construct"
/mol_type="other DNA"
/db_xref="taxon:32630"
<1..>105
/note="(amino acids 31 to 65)"
/codon_start=1
/transl_table=11

CDS

1..105
/transl_table=11

/product="insulin C chain"
/protein_id="CAA24707.1"
/db_xref="GI:929915"
/translation="RREAEDLQVGVELGGPGAGSLQPLALEGSLQKR"

ORIGIN

Alignment Scores:
Pred. No.: 0.0101 Length: 105
Score: 53.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 12 Gaps: 0

US-10-783-095-9 (1-11) x XXHSIN (1-105)

QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnLys 11
DB 70 CAACCTTTGGCTCTCGAGGATCACTTCAAAAG 102

RESULT 4
AX179492
LOCUS AX179492 216 bp DNA linear PAT 31-AUG-2001
DEFINITION Sequence 87 from Patent WO0131037.
ACCESSION AX179492
VERSION AX179492.1 GI:14599120
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.

REFERENCE 1
AUTHORS Ben-Nun,A., Kerlero de Rosbo,N. and Sappier,G.P.
TITLE Synthetic human genes and polypeptides and their use in the treatment of autoimmune diseases
JOURNAL Patent: WO 0131037-A 87 03-MAY-2001;
YEDA RESEARCH AND DEVELOPMENT CO. LTD. (IL)
FEATURES Location/Qualifiers
source 1..216
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
1..207
/note="unnamed protein product; synthetic construct"

CDS

1..207
/codon_start=1
/transl_table=11
/protein_id="CAC43900.1"
/db_xref="GI:14599121"
/translation="MPRLPLALLALWGPDAASHLVREALYVSGRGFFYTPKTR
IEAAGAGSLQPLALEGSLQKIPR"

ORIGIN

Alignment Scores:
Pred. No.: 0.0221 Length: 216
Score: 53.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-783-095-9 (1-11) x AX179492 (1-216)

QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnLys 11
DB 163 CAACCGTTGGCGCTGAGGGTCCCTGCAAAAG 195

RESULT 5

101821
LOCUS 101821 252 bp ss-DNA
DEFINITION Sequence 6 from Patent US 4801536.
ACCESSION 101821
VERSION 101821.1 GI:269797
KEYWORDS Unknown.

linear PAT 21-MAY-1993

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 28, 2004, 11:40:21 ; Search time 143.398 Seconds
(without alignments)
402.682 Million cell updates/sec

Title: US-10-783-095-9

Perfect score: 53

Sequence: 1 QPLALEGSLQK 11

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 413486 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+pin.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10783095/runat_26102004_185425_24797/app_query.fasta_1.796
-DB=N Geneseq_23Sep04 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pfco -NOEM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10783095 -CGCN_1_1241 -runat_26102004_185425_24797 -NCGP
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq_23Sep04:*

1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	53	100.0	66	2	Aax86082 PCR prime
2	53	100.0	87	4	Aaf61522 Human pro
3	53	100.0	159	2	Aat45977 Oligonucle
4	53	100.0	189	2	Aat45973 DNA encod
5	53	100.0	201	2	Aat45974 DNA encod
6	53	100.0	216	5	Aas04793 Synthetic

7	53	100.0	252	1	Aan70988 Sequence
8	53	100.0	258	2	Aat37134 Proinsuli
9	53	100.0	258	10	Adc51568 Human pro
10	53	100.0	258	10	Adf16306 Human alb
11	53	100.0	258	10	Adh21708 Human lon
12	53	100.0	260	12	Adh04119 Proinsuli
13	53	100.0	261	2	Aat17831 Optimised
14	53	100.0	261	2	Aat17830 Human ins
15	53	100.0	264	2	Aat45975 Gene for
16	53	100.0	268	2	AaQ43935 MetTyr hu
17	53	100.0	269	2	Aat80401 Proinsuli
18	53	100.0	272	1	Aan20041 Human pro
19	53	100.0	272	1	Aan60872 Sequence
20	53	100.0	275	2	AaQ37003 Proinsuli
21	53	100.0	278	2	AaQ32256 Synthetic
22	53	100.0	279	2	AaQ83195 Proinsuli
23	53	100.0	281	2	AaQ05663 Human pro
24	53	100.0	281	2	AaQ38310 hpl gene.
25	53	100.0	283	2	AaQ20543 Encodes p
26	53	100.0	287	2	Aat04874 Synthetic
27	53	100.0	289	12	AdL24436 Modified
28	53	100.0	290	12	AdL24441 Modified
29	53	100.0	293	1	Aan90279 Synthetic
30	53	100.0	298	2	AaQ27123 DNA encod
31	53	100.0	298	2	AaQ99025 Chelating
32	53	100.0	304	2	AaQ15230 NcoI-Sali
33	53	100.0	327	6	Abn85338 Oligonucle
34	53	100.0	330	9	Ada09217 Human mRN
35	53	100.0	330	12	AdJ25929 Human pre
36	53	100.0	333	3	AaZ51476 Human ins
37	53	100.0	342	1	Aan40179 Sequence
38	53	100.0	342	1	Aan50152 Sequence
39	53	100.0	342	1	Aan50082 Proinsuli
40	53	100.0	351	3	AaZ59208 MWPs-MWP
41	53	100.0	359	2	Aat75193 Human ins
42	53	100.0	359	2	Aat75652 Human ins
43	53	100.0	359	3	AaC55717 Mutant hu
44	53	100.0	359	4	AaF58802 Human ins
45	53	100.0	360	5	Aas04761 Synthetic

ALIGNMENTS

RESULT 1
AAX86082/c
ID AAX86082 standard; DNA; 66 BP.

XX AC AAX86082;

XX DT 15-SEP-1999 (first entry)

XX DE PCR primer used to produce chimeric genes of the invention.

XX KW Chimeric fusion protein; immunodominant epitope; glutamate decarboxylase;
XX KW GAD; insulin; insulin chain B; human T cell response; GAD65 peptide;
XX KW Type I diabetes; insulin dependent diabetes mellitus; IDDM;
XX KW Stiff Man syndrome; transplant; pancreatic cell; PCR primer; ss.

XX OS Synthetic.

XX OS Homo sapiens.

XX PN WO9932136-A1.

XX PD 01-JUL-1999.

XX PF 23-DEC-1998; 98WO-US027408.

XX PR 23-DEC-1997; 97US-0068648P.

XX PA (ALEX-) ALEXON PHARM INC.

XX PA (MUEL/) MUELLER J P.

XX PA (MATI/) MATIS L A.

PI Mueller JP, Matis LA, Wang Y;
 XX WPI; 1999-405113/34.
 XX
 XX Chimeric fusion proteins comprising immunodominant epitopes of glutamate
 PT decarboxylase and insulin, used to treat patients that have or are
 PT predicted to be at risk of developing type I diabetes.
 XX
 XX Example; Page 62; 67pp; English.
 XX
 XX The specification describes chimeric fusion proteins comprising
 CC immunodominant epitopes of glutamate decarboxylase (GAD) and insulin. The
 CC chimeric fusion proteins comprises insulin chain B and individual peptide
 CC moieties consisting of at least one GAD65 peptide capable of eliciting a
 CC human T cell response, where the insulin chain B and GAD65 peptide are
 CC covalently linked and the chimeric fusion protein is capable of eliciting
 CC a human T cell response to insulin chain B and to each of the at least
 CC one GAD65 peptides. The chimeric fusion proteins are used to treat
 CC patients that are predicted to be at risk of developing Type I diabetes
 CC and those suffering from type I diabetes. The fusion proteins can also be
 CC used to diagnose patients suffering from insulin dependent diabetes
 CC mellitus (IDDM) and/or Stiff Man syndrome. This is also useful for
 CC determining the suitability of patients as recipients of transplants of
 CC pancreatic cells or tissues. The present sequence represents a primer
 CC used to construct genes encoding the chimeric fusion proteins of the
 CC invention
 XX
 SQ Sequence 66 BP; 10 A; 27 C; 21 G; 8 T; 0 U; 0 Other;
 Alignment Scores: Pred. No.: 0.0131 Length: 66
 Score: 53.00 Matches: 11
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 2 Gaps: 0
 US-10-783-095-9 (1-11) x AMX86082 (1-66)
 QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnLys 11
 DB 42 CAGCCCTTGGCGCTGGAGGGGTCCCTGCAGAAG 10
 RESULT 2
 ID AAF61522 standard; DNA; 87 BP.
 AC AAF61522;
 XX
 XX 04-JUL-2001 (first entry)
 DT Human proinsulin C-peptide encoding DNA.
 DE Insulin; proinsulin; C-peptide; human; B-chain; A-chain; diabetes; ds.
 XX Homo sapiens.
 OS DE19947456-A1.
 PN 05-APR-2001.
 PD 02-OCT-1999; 99DE-01047456.
 PF 02-OCT-1999; 99DE-01047456.
 PR (AVET) AVENTIS PHARMA DEUT GMBH.
 XX Habermann P, Ertl J, Meiwes J, Seipke G;
 PI WPI; 2001-267042/28.
 DR P-PSDB; AAB70847.
 XX
 PT New synthetic derivatives of the C-peptide of proinsulin, useful in the

PT preparation of human insulin or insulin analogs in high yield.
 XX Disclosure; Page 7; 12pp; German.
 XX
 XX This invention describes novel synthetic derivatives (I) of the C-peptide
 CC of proinsulin. Such novel derivatives of human insulin or insulin analogs
 CC represented by (I) are of formula Fus-B(1-30)-RDVP-Yn-A(1-21). Fus = the
 CC optionally present fusion component of arbitrary sequence; B(1-30) = the
 CC B-chain of human insulin; Y = amino acid chain having a basic amino acid
 CC at the C-terminal; n = 2-50 (defining the chain length of Y), and A(1-21)
 CC = the A-chain of human insulin; the A- and/or B-chain(s) may be modified
 CC by amino acid exchanges, additions and/or deletions. (I) are
 CC intermediates for human insulin or its analogs, which are used in the
 CC treatment of diabetes. A claimed method of preparing human insulin (or
 CC analogs) involves preparing (I), folding (I) so that disulfide bonds as
 CC in human insulin can form, enzymatically removing the RDVP-Yn part and
 CC Fus (if present) and purifying the product. Insulin (or analogs) can be
 CC prepared in increased yields via (I). Specifically the expression yield
 CC can be increased around 20% compared with that obtained using plasmid
 CC pIN790d and almost 5-fold compared with that obtained using plasmid
 CC pIN302d. The control of enzymatic processing is also improved. The
 CC folding rate is comparable with that obtained using simian proinsulin
 CC encoded by pIN790d. This sequence encodes the human proinsulin C-peptide
 CC which is described in the method of the invention
 XX
 SQ Sequence 87 BP; 10 A; 27 C; 37 G; 13 T; 0 U; 0 Other;
 Alignment Scores: Pred. No.: 0.0181 Length: 87
 Score: 53.00 Matches: 11
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 4 Gaps: 0
 US-10-783-095-9 (1-11) x AAF61522 (1-87)
 QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnLys 11
 DB 52 CAGCCCTTGGCGCTGGAGGGGTCCCTGCAGAAG 84
 RESULT 3
 ID AAT45977 standard; cDNA; 159 BP.
 XX
 AC AAT45977;
 XX
 DT 16-MAR-1997 (first entry)
 DE Oligonucleotide for 3' end of single chain insulin.
 XX Single chain insulin; diabetes mellitus; gene therapy; ss.
 KW Synthetic.
 OS EP741188-A2.
 XX 06-NOV-1996.
 PD 03-MAY-1996; 96EP-00303133.
 PF 05-MAY-1995; 95US-00435503.
 PR 05-MAY-1995; 95US-00435572.
 XX (ELIL) LILLY & CO ELI.
 XX Chance RE, Dimarchi RD, Hoffmann JA, Long HB, Miller AR;
 PI WPI; 1996-487391/49.
 DR Single chain insulin polypeptide(s) - used for treating diabetes.
 XX Example 7; Page 15; 22pp; English.
 XX

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 28, 2004, 11:54:47 ; Search time 32.6024 Seconds
(without alignments)
239.819 Million cell updates/sec

Title: US-10-783-095-9
Perfect score: 53
Sequence: 1 QPLALEGSLQK 11

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q/cgn2_1/USPTO.spool_p/US10783095/runat_26102004_185427_24832/app_query.fasta_1.796
-DB=Issued_Patents NA -QMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-NODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=JUS10783095 @cgn_1_1_213 @runat_26102004_185427_24832 -NCPUS=6 -ICPU=3
-NO MWAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:
5: /cgn2_6/ptodata/1/ina/PTCUS_COMB.seq:
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	53	100.0	68	2	US-08-600-783-18
2	53	100.0	87	4	Sequence 18, Appl
3	53	100.0	258	2	US-09-676-787-1
4	53	100.0	269	2	US-08-600-783-17
5	53	100.0	275	1	Sequence 17, Appl
6	53	100.0	275	1	Sequence 10, Appl
7	53	100.0	275	1	Sequence 12, Appl
8	53	100.0	275	1	Sequence 14, Appl
9	53	100.0	275	1	Sequence 12, Appl
10	53	100.0	276	1	US-08-081-661-14
11	53	100.0	277	1	Sequence 14, Appl
12	53	100.0	278	1	US-07-918-953-16
					Sequence 16, Appl
					Sequence 3, Appl
					Sequence 2, Appl

13	53	100.0	281	1	US-07-764-655D-12
14	53	100.0	281	1	US-07-764-655D-13
15	53	100.0	281	6	5514646-1
16	53	100.0	290	1	US-08-030-731A-41
17	53	100.0	290	1	US-08-030-731A-42
18	53	100.0	298	1	US-07-826-928A-28
19	53	100.0	304	1	US-07-896-551B-12
20	53	100.0	330	4	US-08-472-701-1
21	53	100.0	330	5	PCT-US95-08596-1
22	53	100.0	351	4	US-09-280-030-49
23	53	100.0	359	3	US-08-589-028-3
24	53	100.0	359	3	US-08-784-582-3
25	53	100.0	359	3	US-08-785-271-3
26	53	100.0	390	4	US-09-280-030-48
27	53	100.0	416	3	US-08-945-140-5
28	53	100.0	450	4	US-09-185-852-1
29	53	100.0	450	4	US-09-323-738-1
30	53	100.0	510	1	US-07-918-953-7
31	53	100.0	510	1	US-08-081-661-7
32	53	100.0	515	3	US-08-589-028-1
33	53	100.0	515	3	US-08-784-582-1
34	53	100.0	515	3	US-08-785-271-1
35	53	100.0	2499	4	US-09-775-508C-7
36	53	100.0	4992	4	US-09-015-399-10
37	50	94.3	68	4	US-09-485-286-7
38	50	94.3	74	4	US-09-485-286-9
39	50	94.3	4646	4	US-09-485-286-14
40	48	90.6	51	6	5514646-42
41	48	90.6	51	6	5514646-43
42	47	88.7	598	3	US-09-537-696-11
43	47	88.7	633	3	US-09-537-696-12
44	47	88.7	666	3	US-09-537-696-13
45	39	73.6	99	1	US-07-918-953-3

ALIGNMENTS

RESULT 1

US-08-600-783-18
; Sequence 18 Application US/08600783
; Patent No. 5962267
; GENERAL INFORMATION:
; APPLICANT: SHIN, Hang Cheol
; APPLICANT: CHANG, Seung Gu
; APPLICANT: KIM, Dae Young
; APPLICANT: KIM, Chong Suhli
; TITLE OF INVENTION: Proinsulin Derivative and Process
; TITLE OF INVENTION: for Producing Human Insulin
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SHIN, Hang Cheol
; STREET: Ssangma-Hanshin Apt. 102-1206,
; CITY: Kwangmyung-shi
; STATE: Kyungki-do
; COUNTRY: Republic of Korea
; ZIP: 423-030
; ADDRESSEE: CHANG, Seung Gu
; STREET: Hyundai Apt. 71-203, Apkujong-dong,
; CITY: Kangnam-ku
; STATE: Seoul
; COUNTRY: Republic of Korea
; ZIP: 135-110
; ADDRESSEE: KIM, Dae Young
; STREET: Sosa Jukong Apt. 108-202, Sosa Bon-dong,
; CITY: Bucheon-shi
; STATE: Kyungki-do
; COUNTRY: Republic of Korea
; ZIP: 422-230
; ADDRESSEE: KIM, Chong Suhli
; STREET: Garden Heights Apt. 202-801, #100,

Sequence 12, Appl
Sequence 13, Appl
Patent No. 5514646
Sequence 41, Appl
Sequence 42, Appl
Sequence 28, Appl
Sequence 12, Appl
Sequence 1, Appl
Sequence 49, Appl
Sequence 3, Appl
Sequence 3, Appl
Sequence 48, Appl
Sequence 5, Appl
Sequence 1, Appl
Sequence 7, Appl
Sequence 7, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 7, Appl
Sequence 10, Appl
Sequence 9, Appl
Sequence 14, Appl
Patent No. 5514646
Patent No. 5514646
Sequence 11, Appl
Sequence 12, Appl
Sequence 13, Appl
Sequence 3, Appl

```

/ STREET: Hwangkeum-dong, Soosung-ku
/ CITY: Taegu
/ STATE: Taegu
/ COUNTRY: Republic of Korea
/ ZIP: 706-040
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy Disk, 3.5 inch, 1.44MB storage
/ COMPUTER: IBM PC/AT
/ OPERATING SYSTEM: MS-DOS
/ SOFTWARE: Word Perfect 5.1
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/600,783
/ FILING DATE:
/ CLASSIFICATION: 514
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: KR 95-2751
/ FILING DATE: 15-FEB-1995
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Shahan Islam
/ REGISTRATION NUMBER: 32,507
/ REFERENCE/DOCKET NUMBER:
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (212) 278-1000
/ TELEFAX: (212) 953-7249
/ INFORMATION FOR SEQ ID NO: 18:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 68 bases
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: oligonucleotide
/ ANTI-SENSE: yes
/ US-08-600-783-18

Alignment Scores:
Pred. No.: 0.00108 Length: 68
Score: 53.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-10-783-095-9 (1-11) x US-08-600-783-18 (1-68)
Qy 1 GlnProLeuAlaLeuGluGlySerLeuGlnLys 11
Db 7 CAGCCGTTGGCGCTGGAGGTTCCCTGCAGAAA 39

RESULT 2
US-09-676-787-1
/ Sequence 1, Application US/09676787
/ Patent No. 6534288
/ GENERAL INFORMATION:
/ APPLICANT: HABERMANN, Paul
/ APPLICANT: ERTL, Johann
/ APPLICANT: MEIWEIS, Johannes
/ APPLICANT: SEIPKE, Gerhard
/ TITLE OF INVENTION: C PEPTIDE FOR IMPROVED PREPARATION OF INSULIN AND INSULIN ANALOGS
/ FILE REFERENCE: 38005-0110
/ CURRENT APPLICATION NUMBER: US/09/676,787
/ CURRENT FILING DATE: 2000-10-02
/ PRIOR APPLICATION NUMBER: DE 199 47 456.7
/ PRIOR FILING DATE: 1999-10-02
/ NUMBER OF SEQ ID NOS: 13
/ SOFTWARE: PatentIn version 3.1
/ SEQ-ID NO 1
/ LENGTH: 87
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ US-09-676-787-1

Alignment Scores:
Pred. No.: 0.00146 Length: 87
Score: 53.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-10-783-095-9 (1-11) x US-09-676-787-1 (1-87)
Qy 1 GlnProLeuAlaLeuGluGlySerLeuGlnLys 11
Db 52 CAGCCGTTGGCGCTGGAGGTTCCCTGCAGAA 84

RESULT 3
US-08-600-783-17
/ Sequence 17, Application US/08600783
/ Patent No. 5962267
/ GENERAL INFORMATION:
/ APPLICANT: SHIN, Hang Cheol
/ APPLICANT: CHANG, Seung Gu
/ APPLICANT: KIM, Dae Young
/ APPLICANT: KIM, Chong Suh1
/ TITLE OF INVENTION: Proinsulin Derivative and Process
/ TITLE OF INVENTION: for Producing Human Insulin
/ NUMBER OF SEQUENCES: 36
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: SHIN, Hang Cheol
/ STREET: Ssangma-Hanshin Apt. 102-1206,
/ STREET: #245 Cholsan-dong
/ CITY: Kwangmyung-shi
/ STATE: Kyungki-do
/ COUNTRY: Republic of Korea
/ ZIP: 423-030
/ ADDRESSEE: CHANG, Seung Gu
/ STREET: Hyundai Apt. 71-203, Apjukong-dong,
/ STREET: Kangnam-ku
/ CITY: Seoul
/ STATE: Seoul
/ COUNTRY: Republic of Korea
/ ZIP: 135-110
/ ADDRESSEE: KIM, Dae Young
/ STREET: Sosa Jukong Apt. 108-202, Sosa Bon-dong,
/ STREET: Sosa-ku
/ CITY: Bucheon-shi
/ STATE: Kyungki-do
/ COUNTRY: Republic of Korea
/ ZIP: 422-230
/ ADDRESSEE: KIM, Chong Suh1
/ STREET: Garden Heights Apt. 202-801, #100,
/ STREET: Hwangkeum-dong, Soosung-ku
/ CITY: Taegu
/ STATE: Taegu
/ COUNTRY: Republic of Korea
/ ZIP: 706-040
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy Disk, 3.5 inch, 1.44MB storage
/ COMPUTER: IBM PC/AT
/ OPERATING SYSTEM: MS-DOS
/ SOFTWARE: Word Perfect 5.1
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/600,783
/ FILING DATE:
/ CLASSIFICATION: 514
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: KR 95-2751
/ FILING DATE: 15-FEB-1995
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Shahan Islam
/ REGISTRATION NUMBER: 32,507
/ REFERENCE/DOCKET NUMBER:
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (212) 278-1000
/ TELEFAX: (212) 953-7249
/ INFORMATION FOR SEQ ID NO: 17:

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 28, 2004, 13:47:02 ; Search time 127.494 Seconds
(without alignments)
442.402 Million cell updates/sec

Title: US-10-783-095-9

Perfect score: 53

Sequence: 1. QPLAEGSLQK 11

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 3413475 seqs, 2563800928 residues

Total number of hits satisfying chosen parameters: 6826950

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10783095/runat_26102004_185428_24911/app_query.fasta_1.796
-DB=Published Applications NA -QMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10783095 -CGN_1_1189 -runat_26102004_185428_24911
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:
1: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/prodata/2/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/prodata/2/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/prodata/2/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/prodata/2/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/prodata/2/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/prodata/2/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/prodata/2/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/prodata/2/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/prodata/2/pubpna/US09C_NEW_PUB.seq:
13: /cgn2_6/prodata/2/pubpna/US10A_PUBCOMB.seq:
14: /cgn2_6/prodata/2/pubpna/US10B_PUBCOMB.seq:
15: /cgn2_6/prodata/2/pubpna/US10C_PUBCOMB.seq:
16: /cgn2_6/prodata/2/pubpna/US10D_PUBCOMB.seq:
17: /cgn2_6/prodata/2/pubpna/US10E_PUBCOMB.seq:
18: /cgn2_6/prodata/2/pubpna/US10F_NEW_PUB.seq:
19: /cgn2_6/prodata/2/pubpna/US11_NEW_PUB.seq:
20: /cgn2_6/prodata/2/pubpna/US60_NEW_PUB.seq:
21: /cgn2_6/prodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

§

Result No.	Score	Query Match	Length	DB ID	Description
1	53	100.0	258	15	US-10-383-285-1
2	53	100.0	260	10	US-09-807-742-16
3	53	100.0	260	10	US-09-807-742-17
4	53	100.0	260	17	US-10-746-149-23
5	53	100.0	260	18	US-10-609-019-21
6	53	100.0	330	15	US-10-346-563-1
7	53	100.0	330	16	US-10-321-717-1
8	53	100.0	351	9	US-09-280-030-49
9	53	100.0	350	9	US-09-280-030-48
10	53	100.0	450	9	US-09-804-409A-10
11	53	100.0	450	10	US-09-969-748C-5
12	53	100.0	450	15	US-10-328-813-1
13	53	100.0	450	16	US-10-411-037-43
14	53	100.0	450	16	US-10-411-026-43
15	53	100.0	450	16	US-10-410-962-43
16	53	100.0	450	16	US-10-411-049-43
17	53	100.0	450	17	US-10-410-930-43
18	53	100.0	450	17	US-10-410-997-43
19	53	100.0	450	17	US-10-411-012-43
20	53	100.0	450	17	US-10-287-994-43
21	53	100.0	450	17	US-10-410-913-43
22	53	100.0	1182	9	US-09-919-344-27
23	53	100.0	1626	18	US-10-609-019-27
24	53	100.0	2499	10	US-09-775-508A-1
25	53	100.0	10487	17	US-10-746-149-44
26	53	100.0	10487	18	US-10-609-019-32
27	53	100.0	10512	17	US-10-746-149-42
28	53	100.0	10512	18	US-10-609-019-31
29	53	100.0	10880	17	US-10-746-149-47
30	53	100.0	10895	18	US-10-609-019-41
31	53	100.0	11255	17	US-10-746-149-43
32	53	100.0	11271	18	US-10-609-019-42
33	53	100.0	11332	18	US-10-609-019-43
34	50	94.3	68	16	US-10-430-752A-7
35	50	94.3	74	16	US-10-430-752A-9
36	50	94.3	4646	16	US-10-430-752A-14
37	48	90.6	375	16	US-10-221-677-25
38	48	90.6	1944	16	US-10-221-677-26
39	47	88.7	333	17	US-10-419-539-2
40	47	88.7	598	13	US-10-013-032-11
41	47	88.7	633	13	US-10-013-032-12
42	47	88.7	666	13	US-10-013-032-13
43	47	88.7	1080	17	US-10-419-539-3
44	47	88.7	1217	17	US-10-419-539-4
45	45	84.9	1616	13	US-10-027-632-252243

ALIGNMENTS

RESULT 1

US-10-383-285-1
; Sequence 1, Application US/10383285
; Publication No. US20030200566A1
; GENERAL INFORMATION:
; APPLICANT: STREATHFIELD, STEPHEN
; APPLICANT: HOWARD, JOHN
; TITLE OF INVENTION: PRODUCTION OF INSULIN AND INSULIN-LIKE PROTEINS IN PLANTS
; TITLE OF INVENTION: PLANTS
; FILE REFERENCE: 10040
; CURRENT APPLICATION NUMBER: US/10383,285
; CURRENT FILING DATE: 2003-05-21
; PRIOR APPLICATION NUMBER: 60/362,874
; PRIOR FILING DATE: 2002-03-08
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 258
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-383-285-1

Alignment Scores:
 Pred. No.: 0.0152 Length: 258
 Score: 53.00 Matches: 11
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 15 Gaps: 0

US-10-783-095-9 (1-11) x US-10-383-285-1 (1-258)

QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnLys 11
 |||||
 Db 160 CAGCCCTTGCCCTCGAGGGCTCCCTCCAGAAG 192

RESULT 2

US-09-807-742-16
 ; Sequence 16, Application US/09807742
 ; Publication No. US20030204864A1
 ; GENERAL INFORMATION:
 ; APPLICANT: DANIELL, HENRY
 ; TITLE OF INVENTION: PRODUCTION OF PHARMACEUTICAL PROTEINS IN TRANSGENIC
 ; FILE OF INVENTION: PLASTIDS
 ; FILE REFERENCE: 1465-PCT-US-00
 ; CURRENT APPLICATION NUMBER: US/09/807,742
 ; CURRENT FILING DATE: 2001-04-18
 ; PRIOR APPLICATION NUMBER: PCT/US01/06288
 ; PRIOR FILING DATE: 2001-02-28
 ; NUMBER OF SEQ ID NOS: 19
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 16
 ; LENGTH: 260
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-807-742-16

Alignment Scores:
 Pred. No.: 0.0153 Length: 260
 Score: 53.00 Matches: 11
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 10 Gaps: 0

US-10-783-095-9 (1-11) x US-09-807-742-16 (1-260)

QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnLys 11
 |||||
 Db 160 CAGCCCTTGCCCTCGAGGGCTCCCTCCAGAAG 192

RESULT 3

US-09-807-742-17
 ; Sequence 17, Application US/09807742
 ; Publication No. US20030204864A1
 ; GENERAL INFORMATION:
 ; APPLICANT: DANIELL, HENRY
 ; TITLE OF INVENTION: PRODUCTION OF PHARMACEUTICAL PROTEINS IN TRANSGENIC
 ; FILE OF INVENTION: PLASTIDS
 ; FILE REFERENCE: 1465-PCT-US-00
 ; CURRENT APPLICATION NUMBER: US/09/807,742
 ; CURRENT FILING DATE: 2001-04-18
 ; PRIOR APPLICATION NUMBER: PCT/US01/06288
 ; PRIOR FILING DATE: 2001-02-28
 ; NUMBER OF SEQ ID NOS: 19
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 17
 ; LENGTH: 260
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Chloroplast
 ; OTHER INFORMATION: modified proinsulin sequence
 US-09-807-742-17

Alignment Scores:
 Pred. No.: 0.0153 Length: 260
 Score: 53.00 Matches: 11
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 10 Gaps: 0

US-10-783-095-9 (1-11) x US-09-807-742-17 (1-260)

QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnLys 11
 |||||
 Db 160 CACCTTTAGCTTTAGAGGTTCTTTACAAAAA 192

RESULT 4

US-10-746-149-23
 ; Sequence 23, Application US/10746149
 ; Publication No. US20040172667A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cooper, Richard K.
 ; APPLICANT: Fioretti, William C.
 ; APPLICANT: Cadd, Gary G.
 ; TITLE OF INVENTION: Administration of Transposon-Based Vectors to Reproductive Organs
 ; FILE REFERENCE: 51687-0280 (51687-294923)
 ; CURRENT APPLICATION NUMBER: US/10/746,149
 ; CURRENT FILING DATE: 2003-12-24
 ; PRIOR APPLICATION NUMBER: US 60/392,415
 ; PRIOR FILING DATE: 2002-06-26
 ; PRIOR APPLICATION NUMBER: US 60/441,381
 ; PRIOR FILING DATE: 2003-01-21
 ; PRIOR APPLICATION NUMBER: US 60/441,447
 ; PRIOR FILING DATE: 2003-01-21
 ; PRIOR APPLICATION NUMBER: US 60/441,405
 ; PRIOR FILING DATE: 2003-01-21
 ; PRIOR APPLICATION NUMBER: US 60/441,502
 ; PRIOR FILING DATE: 2003-01-21
 ; PRIOR APPLICATION NUMBER: US 60/441,392
 ; PRIOR FILING DATE: 2003-01-21
 ; PRIOR APPLICATION NUMBER: US 60/441,377
 ; PRIOR FILING DATE: 2003-01-21
 ; PRIOR APPLICATION NUMBER: US 10/509,019
 ; PRIOR FILING DATE: 2003-06-26
 ; NUMBER OF SEQ ID NOS: 52
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 23
 ; LENGTH: 260
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Synthetic
 US-10-746-149-23

Alignment Scores:
 Pred. No.: 0.0153 Length: 260
 Score: 53.00 Matches: 11
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 17 Gaps: 0

US-10-783-095-9 (1-11) x US-10-746-149-23 (1-260)

QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnLys 11
 |||||
 Db 160 CAGCCCTTGCCCTCGAGGGCTCCCTCCAGAAG 192

RESULT 5

US-10-609-019-21
 ; Sequence 21, Application US/10609019
 ; Publication No. US20040197910A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cooper, Richard K.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2004, 10:15:44 ; Search time 8.87952 Seconds
(without alignments)
119.194 Million cell updates/sec

Title: US-10-783-095-9
Perfect score: 53
Sequence: 1 QPLAEGSLQK 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_79:*
1: Pir1:*
2: Pir2:*
3: Pir3:*
4: Pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	53	100.0	110	1	IPHU
2	53	100.0	110	2	insulin precursor
3	53	100.0	110	2	insulin precursor
4	53	100.0	110	2	insulin precursor
5	50	94.3	110	1	IPDG
6	41	77.4	110	1	IPGP
7	41	77.4	128	2	S77068
8	36	67.9	156	2	PC4141
9	36	67.9	210	2	T25703
10	36	67.9	374	2	T13262
11	36	67.9	374	2	H86750
12	36	67.9	374	2	A47085
13	36	67.9	597	2	H71809
14	36	67.9	597	2	D64711
15	36	67.9	598	2	C81281
16	36	67.9	608	2	C97575
17	36	67.9	608	2	AC2796
18	36	67.9	609	2	AF2379
19	35	66.0	1084	2	F95141
20	35	66.0	1084	2	D98009
21	34	64.2	110	1	INRB
22	34	64.2	388	2	C87221
23	34	64.2	456	2	T40386
24	34	64.2	541	2	A12238
25	34	64.2	584	2	B90281
26	34	64.2	591	2	G90181
27	34	64.2	592	2	D70327
28	34	64.2	605	2	S18561
29	34	64.2	608	1	S01040

RESULT 1

IPHU

insulin precursor [validated] - human

N;Alternate names: preproinsulin

C;Species: Homo sapiens (man)

C;Date: 23-Oct-1981 #sequence revision 23-Oct-1981 #text change 09-Jul-2004

C;Accession: A93222; A94253; A93216; A94251; A93144; A92075; A91186; I58114; A01579; S58

R;Bell, G.I.; Pictet, R.L.; Rutter, W.J.; Cordell, B.; Fischer, E.; Goodman, H.M.

Nature 284, 26-32, 1980

A;Title: Sequence of the human insulin gene.

A;Reference number: A93222; MUID:90120725; PMID:6243748

A;Accession: A93222

A;Molecule type: DNA

A;Residues: 1-110 <BEL>

A;Cross-references: UNIPROT:P01308; GB:J00265; NID:G186429; PIDN:AAA59172.1; PID:G386828

R;Ullrich, A.; Dull, T.J.; Gray, A.; Brosius, J.; Sures, I.

Science 209, 612-615, 1980

A;Title: Genetic variation in the human insulin gene.

A;Reference number: A94253; MUID:80236313; PMID:6248962

A;Accession: A94253

A;Molecule type: DNA

A;Residues: 1-110 <ULL>

A;Cross-references: GB:J00265; NID:G186429; PIDN:AAA59172.1; PID:G386828

R;Bell, G.I.; Swain, W.F.; Pictet, R.; Cordell, B.; Goodman, H.M.; Rutter, W.J.

Nature 282, 525-527, 1979

A;Title: Nucleotide sequence of a cDNA clone encoding human preproinsulin.

A;Reference number: A93216; MUID:80054779; PMID:503234

A;Accession: A93216

A;Molecule type: mRNA

A;Residues: 1-110 <BEL2>

A;Cross-references: GB:J00265; NID:G186429; PIDN:AAA59172.1; PID:G386828

R;Sures, I.; Goeddel, D.V.; Gray, A.; Ullrich, A.

Science 208, 57-59, 1980

A;Title: Nucleotide sequence of human preproinsulin complementary DNA.

A;Reference number: A94251; MUID:80147417; PMID:6927840

A;Accession: A94251

A;Molecule type: mRNA

A;Residues: 1-110 <SUR>

A;Cross-references: GB:J00265; NID:G186429; PIDN:AAA59172.1; PID:G386828

R;Nicol, D.S.H.W.; Smith, L.F.

Nature 187, 483-485, 1960

A;Title: Amino-acid sequence of human insulin.

A;Reference number: A93144

A;Accession: A93144

A;Molecule type: protein

A;Residues: 25-54;90-110 <NIC>

R;Oyer, P.E.; Cho, S.; Peterson, J.D.; Steiner, D.F.

J. Biol. Chem. 246, 1375-1386, 1971

A;Title: Studies on human proinsulin. Isolation and amino acid sequence of the human pan

A;Reference number: A92075; MUID:V116410; PMID:5101771

A;Accession: A92075

A;Molecule type: protein

A:Residues: 57-87 <OYE>
 R:Ko, A.; Smyth, D.G.; Markussen, J.; Sundby, F.
 Eur. J. Biochem. 20, 190-199, 1971
 A:Title: Amino acid sequence of the C-peptide of human proinsulin.
 A:Reference number: A91186; MUID:71257722; PMID:5560404
 A:Accession: A91186
 A:Molecule type: protein
 A:Residues: 57-87 <SKOA>
 R:Lucassen, A.M.; Jullier, C.; Beressi, J.P.; Boitard, C.; Froguel, P.; Lathrop, M.; Bell
 Nature Genet. 4, 305-310, 1993
 A:Title: Susceptibility to insulin dependent diabetes mellitus maps to a 4.1 kb segment
 A:Reference number: I58114; MUID:93364428; PMID:8358440
 A:Accession: I58114
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-59,63-110 <RES>
 A:Cross-references: GB:L15440; NID:9307071; PIDN:AAA59179.1; PID:9307072
 R:Sieber, P.; Kamber, B.; Hartmann, A.; Joehl, A.; Riniker, B.; Rittel, W.
 Helv. Chim. Acta 57, 2617-2621, 1974
 A:Title: Totalsynthese von Humaninsulin unter gezielter Bildung der Disulfidbindungen.
 A:Reference number: A91636; MUID:75077277; PMID:4443293
 A:Contents: annotation; synthesis
 A:Note: disulfide-bonded human insulin was synthesized; the synthetic hormone was identical
 A:Note: article in German with English abstract
 R:Naithani, V.K.
 Hoppe-Seyler's Z. Physiol. Chem. 354, 659-672, 1973
 A:Title: The synthesis of C-peptide of human proinsulin.
 A:Reference number: A91658; MUID:75040007; PMID:4803504
 A:Contents: annotation; synthesis of residues 57-87
 R:Geiger, R.; Jaeger, G.; Koenig, W.
 Chem. Ber. 106, 2347-2352, 1973
 A:Title: Synthesis of the complete sequence of human proinsulin C-peptide and its [Glu-9
 A:Reference number: A50914
 A:Contents: annotation; synthesis of residues 57-87
 R:Kaufmann, J.E.; Irminger, J.C.; Halban, P.A.
 Biochem. J. 310, 859-874, 1995
 A:Title: Sequence requirements for proinsulin processing at the B-chain/C-peptide junction
 A:Reference number: S58661; MUID:96013185; PMID:7575420
 A:Contents: annotation; site-directed mutagenesis study of proteolytic processing
 C:Genetics:
 A:Gene: GDB:INS
 A:Cross-references: GDB:119349; OMIM:176730
 A:Map position: 11p15.5-11p15.5
 A:Introns: 63/1
 C:Superfamily: insulin
 C:Keywords: hormone; pancreas
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-54/Domain: insulin chain B #status experimental <BCH>
 F:725-54,90-110/Product: insulin #status experimental <MAT>
 F:57-87/Domain: connecting C peptide #status experimental <CPEP>
 F:90-110/Domain: insulin chain A #status experimental <ACH>
 F:31-96,43-109,95-100/Disulfide bonds: #status experimental
 Query Match 100.0%; Score 53; DB 1; Length 110;
 Best Local Similarity 100.0%; Pred. No. 0.0011;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 QPAALEGSLOK 11
 DB 78 QPAALEGSLOK 88
 RESULT 2
 B42179
 insulin precursor - green monkey
 C:Species: Cercopithecus aethiops (green monkey, grivet)
 C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
 C:Accession: B42179; A05232; S16494; S22056
 R:Seino, S.; Bell, G.I.; Li, W.H.
 Mol. Biol. Evol. 9, 193-203, 1992
 A:Title: Sequences of primate insulin genes support the hypothesis of a slower rate of m
 A:Reference number: A42179; MUID:92219953; PMID:1560757
 A:Accession: B42179
 A:Molecule type: DNA
 A:Residues: 1-110 <SEI>
 A:Cross-references: UNIPROT:P30407; EMBL:X61092; NID:922808; PIDN:CAA43405.1; PID:922809
 A:Note: sequence extracted from NCBI backbone (NCBIN:95185, NCBIP:95194)
 R:Peterson, J.D.; Nehrlich, S.; Oyer, P.E.; Steiner, D.F.
 J. Biol. Chem. 247, 4866-4871, 1972
 A:Title: Determination of the amino acid sequence of the monkey, sheep, and dog proinsulin;
 A:Reference number: A92111; MUID:72258016; PMID:4626369
 A:Accession: A05232
 A:Molecule type: protein
 A:Residues: 57-87 <PEI>
 C:Genetics:
 A:Introns: 63/1
 C:Superfamily: insulin
 C:Keywords: hormone; pancreas
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-54/Domain: insulin chain B #status predicted <BCH>
 F:25-54,90-110/Product: insulin #status predicted <MAT>
 F:57-87/Domain: connecting peptide #status experimental <CPEP>
 F:90-110/Domain: insulin chain A #status predicted <ACH>
 F:31-96,43-109,95-100/Disulfide bonds: #status predicted
 Query Match 100.0%; Score 53; DB 2; Length 110;
 Best Local Similarity 100.0%; Pred. No. 0.0011;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 QPAALEGSLOK 11
 DB 78 QPAALEGSLOK 88
 RESULT 4
 A42179
 insulin precursor - chimpanzee
 C:Species: Pan troglodytes (chimpanzee)
 C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
 C:Accession: A42179; S22058
 R:Seino, S.; Bell, G.I.; Li, W.H.
 Mol. Biol. Evol. 9, 193-203, 1992
 A:Title: Sequences of primate insulin genes support the hypothesis of a slower rate of m
 A:Reference number: A42179; MUID:92219953; PMID:1560757
 A:Accession: A42179

clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
Possible reversed clone: similarity on wrong strand
Seq primer: -40RP from Gibco.

FEATURES

source

1. .118
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5779548"
/tissue_type="insulinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Human insulinoma"
/note="Organ: pancreas; Vector: pBluescript SK-; Site 1:
XhoI; Site 2: EcoRI; Constructed with lambda ZAPII system
(Stratagene) by Dr. J. Ferrer, in vivo mass-excised to
pBluescript SK- by Dr. H. Inoue following the Washington
University protocol
(http://genome.wustl.edu/est/lambda_protocol.shtml).
Please contact Hiroshi Inoue, MD/PhD for further
information on this library (Metabolism Division, Permutt
Laboratory, Washington University School of Medicine, Box
8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this
is a Washington University Pancreas EST project library."

ORIGIN

Alignment Scores:
Pred. No.: 0.324 Length: 118
Score: 53.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 5 Gaps: 0

US-10-783-095-9 (1-11) x BQ267254 (1-118)

QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnLys 11
DB 113 CAGCCCTTGGCCCTGGAGGGTCCCTGCAGAG 81

RESULT 2

BQ272508 118 bp mRNA linear EST 15-JUL-2003
LOCUS ik01b07.x1 Human insulinoma Homo sapiens cDNA clone IMAGE:5779548
DEFINITION 3' similar to SW:INS_HUMAN P01308 INSULIN PRECURSOR. [1] ; mRNA
sequence.

ACCESSION

BQ272508

VERSION BQ272508.1 GI:20497577

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 118)

Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,

Lenishka,I., Scarce,M., Brestelli,J., Gradwohl,G., Clifton,S.,

Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,

Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J.,

Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagarisvili,R.,

Williams,T., Jackson,Y. and Bowers,Y.

Endocrine Pancreas Consortium

Unpublished (2000)

Other ESTs: ik01b07.y1

Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue

Endocrine Pancreas Consortium

Harvard University, Howard Hughes Medical Institute

Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,

MA 02138

Tel: 617-495-1812

Fax: 617-495-8557

Email: dmelton@biohp.harvard.edu

Library was constructed by Dr. J. Ferrer In vivo mass-excised to

pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington

University Genome Sequencing Center For information on obtaining a

clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
Possible reversed clone: similarity on wrong strand
Seq primer: -40UP from Gibco.

FEATURES

source

1. .118
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5779548"
/tissue_type="insulinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Human insulinoma"
/note="Organ: pancreas; Vector: pBluescript SK-; Site 1:
XhoI; Site 2: EcoRI; Constructed with lambda ZAPII system
(Stratagene) by Dr. J. Ferrer, in vivo mass-excised to
pBluescript SK- by Dr. H. Inoue following the Washington
University protocol
(http://genome.wustl.edu/est/lambda_protocol.shtml).
Please contact Hiroshi Inoue, MD/PhD for further
information on this library (Metabolism Division, Permutt
Laboratory, Washington University School of Medicine, Box
8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this
is a Washington University Pancreas EST project library."

ORIGIN

Alignment Scores:
Pred. No.: 0.324 Length: 118
Score: 53.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 5 Gaps: 0

US-10-783-095-9 (1-11) x BQ272508 (1-118)

QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnLys 11
DB 5 CAGCCCTTGGCCCTGGAGGGTCCCTGCAGAG 37

RESULT 3

BQ073408 118 bp mRNA linear EST 27-AUG-2002
LOCUS im35e06.x1 Human insulinoma Homo sapiens cDNA clone IMAGE:6036971
DEFINITION 3' similar to SW:INS_HUMAN P01308 INSULIN PRECURSOR. [1] ; mRNA
sequence.

ACCESSION

BQ073408

VERSION BQ073408.1 GI:22514597

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 118)

Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,

Lenishka,I., Scarce,M., Brestelli,J., Gradwohl,G., Clifton,S.,

Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,

Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J.,

Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagarisvili,R.,

Williams,T., Jackson,Y. and Bowers,Y.

Endocrine Pancreas Consortium

Unpublished (2000)

Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue

Endocrine Pancreas Consortium

Harvard University, Howard Hughes Medical Institute

Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,

MA 02138

Tel: 617-495-1812

Fax: 617-495-8557

Email: dmelton@biohp.harvard.edu

Library was constructed by Dr. J. Ferrer In vivo mass-excised to

pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington

University Genome Sequencing Center For information on obtaining a

clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)

Q38608 lactococcus
Q9b016 bacterioph
Q9cgt4 lactococcus
Q7mau3 wolinnella s
Q9zj94 h glucosami
Q26060 h glucosami
Q9pmt4 c glucosami
Q7x5j2 helicobacte
Q8ueh1 a glucosami
P59362 b glucosami
P94323 bradyrhizob
Q6ssy6 uncultured
Q72sf2 desulfovibr
Aar38436 unculture

32 36 67.9 374 2 Q38608
33 36 67.9 374 2 Q9b016
34 36 67.9 374 2 Q9cgt4
35 36 67.9 595 2 Q7mau3
36 36 67.9 596 1 GLMS_HELPJ
37 36 67.9 596 1 GLMS_HELPY
38 36 67.9 597 1 GLMS_CANOE
39 36 67.9 597 2 Q7x5j2
40 36 67.9 607 1 GLMS_AGR75
41 36 67.9 607 1 GLMS_BRAJA
42 36 67.9 607 1 NODM_BRAJA
43 36 67.9 607 2 Q6SEV6
44 36 67.9 607 2 Q72sf2
45 36 67.9 607 2 Aar38436

ALIGNMENTS

RESULT 1
Q8HZ80 PRELIMINARY; PRT; 65 AA.
ID Q8HZ80
AC Q8HZ80;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Insulin (fragment).
OS Pongo pygmaeus (Orangutan).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
OX NCBI_TaxID=9600;
RN [1]
RP SEQUENCE FROM N.A.
RA O'Huigin C., Tichy H., Klein J.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY092024; AAM76641.1;
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0005179; F:hormone activity; IEA.
DR GO; GO:0007582; P:physiological process; IEA.
DR InterPro; IPR004825; Ins/IGF/relax.
DR Pfam; PF00049; Insulin; 1.
FT NON_TER 1 65
FT NON_TER 65 65
SQ SEQUENCE 65 AA; 6920 MW; B772017FD8BCABEA CRC64;

Query Match 100.0%; Score 53; DB 2; Length 65;
Best Local Similarity 100.0%; Pred. No. 0.006;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Cq 1 QPLALEGSLOK 11
Db 48 QPLALEGSLOK 58

RESULT 2
Q8HZ81 PRELIMINARY; PRT; 65 AA.
ID Q8HZ81
AC Q8HZ81;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Insulin (fragment).
OS Gorilla gorilla (gorilla).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
OX NCBI_TaxID=9593;
RN [1]
RP SEQUENCE FROM N.A.
RA O'Huigin C., Tichy H., Klein J.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -|- SUBCELLULAR LOCATION: Secreted (By similarity).
CC -|- SIMILARITY: Belongs to the insulin family.
DR EMBL; AY092023; AAM76640.1; -.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2004, 06:48:48 ; Search time 38.9639 Seconds
(without alignments)
182.436 Million cell updates/sec

Title: US-10-783-095-9
Perfect score: 53
Sequence: 1 QPLALEGSLOK 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_02:
1: uniprot_sprot:
2: uniprot_trembl:

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	53	100.0	65	Q8HZ80	Q8hz80 pongo pygma
2	53	100.0	65	Q8HZ81	Q8hz81 gorilla gor
3	53	100.0	110	INS_CERAE	P30407 cercopithec
4	53	100.0	110	INS_HUMAN	P01308 homo sapien
5	53	100.0	110	INS_MACFA	P30406 macaca fasc
6	53	100.0	110	INS_PANTR	P30410 pan troglod
7	53	100.0	110	INS_PONPY	Q8bxv2 pongo pygma
8	53	100.0	110	Q6YK33	Q6yk33 gorilla gor
9	53	100.0	110	AAP35454	Aap35454 homo sapi
10	53	100.0	110	AAN06935	Aan06935 gorilla g
11	53	100.0	110	AAN39451	Aan39451 homo sapi
12	50	94.3	110	INS_CANFA	P01321 canis fami
13	41	77.4	110	INS_CAVPO	P01329 cavia porce
14	41	77.4	110	INS_SPSTR	Q91xi3 spermophilu
15	41	77.4	128	Q55981	Q55981 synchocyst
16	40	75.5	1164	Q72CD0	Q72cd0 desulfovibr
17	40	75.5	1164	AAS95831	Aas95831 desulfovi
18	39	73.6	188	Q88AB3	Q88ab3 pseudomonas
19	38	71.7	670	Q96E24	Q96e24 homo sapien
20	38	71.7	675	Q8BT66	Q8bt66 mus musculu
21	38	71.7	679	Q8TF50	Q8tf50 homo sapien
22	37	69.8	117	GHRL_CANFA	Q9bzf8 canis fami
23	37	69.8	117	BAC75929	Bac75929 canis fam
24	36	67.9	76	Q9BYA1	Q9byal homo sapien
25	36	67.9	110	Q8WNW6	Q8wnw6 felis silve
26	36	67.9	156	GLMS_SPHYA	Q56206 s glucosami
27	36	67.9	210	P91262	P91262 caenorhabdi
28	36	67.9	358	Q7V7S5	Q7v7s5 prochloroco
29	36	67.9	374	Q38088	Q38088 bacterioph
30	36	67.9	374	Q38313	Q38313 bacterioph
31	36	67.9	374	Q38325	Q38325 lactococcus

```

DR GO: GO:0005576; C:extracellular; IEA.
DR GO: GO:0005179; F:hormone activity; IEA.
DR GO: GO:0007582; P:physiological process; IEA.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR SMART: SM00078; IIGF; 1.
KW Insulin family.
FT NON_TER 1
FT NON_TER 65
SQ SEQUENCE 65 AA; 6920 MW; B772017FD8BCABEA CRC64;

Query Match 100.0%; Score 53; DB 2; Length 65;
Best Local Similarity 100.0%; Pred. No. 0.006;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPLALEGSLQK 11
DB 48 QPLALEGSLQK 58

RESULT 3
INS_CERAE
ID INS_CERAE STANDARD; PRT; 110 AA.
AC P30407; P01309;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Insulin precursor.
GN Name=INS;
OS Cercopithecus aethiops (Green monkey) (Grivet).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Cercopithecus.
OX NCBI_TaxID=9534;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92219953; PubMed=1560757;
RA Seino S., Bell G.I., Li W.;
RT "Sequences of primate insulin genes support the hypothesis of a slower
rate of molecular evolution in humans and apes than in monkeys.";
RL Mol. Biol. Evol. 9:193-203(1992).
RN [2]
RP SEQUENCE OF 57-87.
EX MEDLINE=72258016; PubMed=4626369;
RA Peterson J.D., Nehrllich S., Oyer P.E., Steiner D.F.;
RT "Determination of the amino acid sequence of the monkey, sheep, and
dog proinsulin C-peptides by a semi-micro Edman degradation
procedure.";
RL J. Biol. Chem. 247:4866-4871(1972).
CC -!- FUNCTION: Insulin decreases blood glucose concentration. It
increases cell permeability to monosaccharides, amino acids and
fatty acids. It accelerates glycolysis, the pentose phosphate
cycle, and glycogen synthesis in liver.
CC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
disulfide bonds.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Belongs to the insulin family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC EMBL; X61092; CAA43405.1; -
CC PIR; B42179; B42179.
CC HSP; P01308; 1A10.
CC InterPro: IPR004825; Ins/IGF/relax.
CC Pfam: PF00049; Insulin; 1.
CC PRINTS; PR00277; INSULIN.
CC SMART; SM00078; IIGF; 1.

DR PROSITE; PS00262; INSULIN; 1.
KW Direct protein sequencing; Glucose metabolism; Hormone;
KW Insulin family; Signal.
FT SIGNAL 1 24
FT CHAIN 25 54 Insulin B chain.
FT PROPEP 57 87 C peptide.
FT CHAIN 90 110 Insulin A chain.
FT DISULFID 31 96 Interchain.
FT DISULFID 43 109 Interchain.
FT DISULFID 95 100
SQ SEQUENCE 110 AA; 12019 MW; 95A1F54BE7B247F9 CRC64;

Query Match 100.0%; Score 53; DB 1; Length 110;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPLALEGSLQK 11
DB 78 QPLALEGSLQK 88

RESULT 4
INS_HUMAN
ID INS_HUMAN STANDARD; PRT; 110 AA.
AC P01308;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Insulin precursor.
GN Name=INS;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=80120725; PubMed=6243748;
RA Bell G.I., Pictet R.L., Rutter W.J., Cordell B., Tischler E.,
RA Goodman H.M.;
RT "Sequence of the human insulin gene.";
RL Nature 284:26-32(1980).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=80236313; PubMed=6248962;
RA Ullrich A., Dull T.J., Gray A., Brosius J., Sures I.;
RT "Genetic variation in the human insulin gene.";
RL Science 209:612-615(1980).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=80054779; PubMed=503234;
RA Bell G.I., Swain W.F., Pictet R.L., Cordell B., Goodman H.M.,
RA Rutter W.J.;
RT "Nucleotide sequence of a cDNA clone encoding human preproinsulin.";
RL Nature 282:525-527(1979).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=8017417; PubMed=6927840;
RA Sures I., Goeddel D.V., Gray A., Ullrich A.;
RT "Nucleotide sequence of human preproinsulin complementary DNA.";
RL Science 208:57-59(1980).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=93364428; PubMed=8358440;
RA Lucassen A.M., Bell J.I., Julier C., Lathrop M.;
RT "Susceptibility to insulin dependent diabetes mellitus maps to a 4.1
kb segment of DNA spanning the insulin gene and associated VNTR.";
RL Nat. Genet. 4:305-310(1993).
RN [6]
RP SEQUENCE FROM N.A.
RX TISSUE=Pancreas;
RC MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2004, 10:55:55 ; Search time 66.506 Seconds
(without alignments)
129.454 Million cell updates/sec

Title: US-10-783-095-12

Perfect score: 24
Sequence: 1 KLVESPSRSDYINASPIIHDP 24

Scoring table: OLIGO
Gapcp 60.0 , Gapext 60.0

Searched: 2002273 seqs, 358729299 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_23sep04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	* Match	Length	ID	Description
1	24	100.0	294	2	Aay33360 Human isl
2	24	100.0	294	2	Aay49854 Human pan
3	24	100.0	294	2	Aaw80486 Islet cel
4	24	100.0	294	3	Aay49329 Pancreat
5	24	100.0	379	3	Aay59351 Tyrosine
6	24	100.0	548	2	Aay33361 Human isl
7	24	100.0	548	2	Aay49855 Human pan
8	24	100.0	548	2	Aaw80487 Islet cel
9	24	100.0	548	3	Aay49330 Pancreat
10	24	100.0	548	6	ABU04788 Human exp
11	24	100.0	548	6	ABU04790 Human exp
12	24	100.0	548	6	ABU04792 Human exp
13	24	100.0	548	6	ABU04791 Human exp
14	24	100.0	591	6	ABU04783 Human exp
15	24	100.0	591	6	ABU04785 Human exp
16	24	100.0	591	6	ABU04795 Human exp
17	24	100.0	916	2	Aar45779 PTPase PT
18	24	100.0	916	6	ABU04786 Human exp
19	24	100.0	923	7	Ades6165 Rat Prote
20	24	100.0	923	7	ADD45276 Rat Prote
21	24	100.0	934	5	ABB06112 Human NS
22	24	100.0	961	2	Aar49039 PTPase PT
23	24	100.0	961	6	ABU04787 Human exp
24	24	100.0	978	8	ADO60054 CRH signa
25	24	100.0	979	2	Aay06606 Human isl

26	24	100.0	979	4	AAG80205 Human aut
27	24	100.0	979	6	ABU04793 Human exp
28	24	100.0	979	6	ABU04784 Human exp
29	24	100.0	979	6	ABU04789 Human exp
30	24	100.0	979	6	ABU04794 Human exp
31	24	100.0	979	7	Ades7230 Human Pro
32	24	100.0	979	7	Ades6167 Human Pro
33	24	100.0	979	7	Ades7226 Human Pro
34	24	100.0	979	7	ADD45278 Human Pro
35	24	100.0	979	7	Ades7218 Human Pro
36	24	100.0	979	7	Ades7222 Human Pro
37	24	100.0	996	5	ABB57041 Mouse isc
38	22	91.7	22	3	AAB30279 CD4+ T-ce
39	19	79.2	19	3	AAB30280 CD4+ T-ce
40	17	70.8	17	3	AAB30278 CD4+ T-ce
41	12	50.0	12	3	AAB30276 CD4+ T-ce
42	12	50.0	12	3	AAB30277 CD4+ T-ce
43	11	45.8	11	3	AAB30291 CD4+ T-ce
44	9	37.5	16	3	AAY59323 Tyrosine
45	9	37.5	298	2	AAW35300 Human isl

ALIGNMENTS

RESULT 1

RAY33360
ID AAY33360 standard; protein; 294 AA.

XX AC AAY33360;

XX AC AAY33360;

DT 29-NOV-1999 (first entry)

DE Human islet cell antigen clone ICA-512 protein fragment.

XX Islet cell antigen; ICA; human; pancreas; antibody; IDDM; immunoassay;
insulin-dependent type I diabetes mellitus; diagnosis; immunoglobulin;
T-cell; B-cell; islet cell destruction; immunoreactivity; disease;
immune therapy; stimulating antigen; screening.

XX Homo sapiens.

XX US9595345-A.

XX 21-SEP-1999.

XX 06-JUN-1995; 95US-00468576.

XX 17-FEB-1989; 89US-00312543.

XX 04-DEC-1989; 89US-00441703.

XX 14-JUN-1991; 91US-00715181.

XX 08-JUN-1992; 92US-00872646.

XX 05-MAY-1994; 94US-00239276.

XX (FARB) BAYER CORP FORMERLY MOLECULAR DIAGNOSTIC.

XX Rabin DJ;

XX WPI; 1999-560494/47.

XX N-PSDB; AAZ09922.

XX Isolated DNA encoding antigens reactive with serum from diabetics useful
to develop products for the diagnosis and monitoring of insulin-dependent
diabetes mellitus.

XX Example; Col 85-87; 63pp; English.

XX This invention describes novel isolated and purified DNAs encoding
pancreatic islet cell antigens reactive with islet cell antibodies found
in sera of patients with insulin-dependent type I diabetes mellitus
(IDDM). The polypeptides of the invention are useful as immunoassay
reagents in the presymptomatic diagnosis of insulin-dependent type I
diabetes mellitus (IDDM). The ability of the polypeptides to bind the

CC antibody binding site on islet cell antibodies (ICAs) also confers
 CC utility in the binding or blocking of human immunoglobulin, T-cells, or B
 CC -cells involved in IDDM. The products can also be used to study the
 CC biological mechanisms involved in islet cell destruction and the
 CC appearance of ICA. The immunoreactivity profile with different antigens
 CC can provide diagnostically significant information concerning the nature
 CC of the disease, e.g. subtypes, the state of the disease, the proximity to
 CC onset of the disease, and the efficacy of therapy, e.g. immune therapy.
 CC The antigens can serve as stimulating antigens for T-cell culture,
 CC permitting significantly improved T-cell cloning, identification, and
 CC growth. The availability of large quantities of pure antigen enables the
 CC development of highly sensitive and specific immunoassays which can be
 CC used to screen the general population of presymptomatic IDDM or a
 CC predisposition to develop IDDM. This sequence represents the human islet
 CC cell antigen from clone ICA-512
 XX
 SQ Sequence 294 AA;

Query Match 100.0%; Score 24; DB 2; Length 294;
 Best Local Similarity 100.0%; Pred. No. 4e-16;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVKVESSPSRSDYINASPIIEHDP 24
 |||||
 Db 110 KLVKVESSPSRSDYINASPIIEHDP 133

RESULT 2
 AA49854
 ID AAY49854 standard; protein; 294 AA.
 AC AAY49854;
 XX
 DT 21-JAN-2000 (first entry)
 XX
 DE Human pancreatic islet cell antigen ICA-512 ATCC 40706 protein.
 KW Human; pancreatic islet cell antigen; ICA clone; IDDM; diabetes;
 KW insulin dependent type I diabetes mellitus; diagnosis; detection;
 KW immunoglobulin; T-cell; B-cell; antibody binding.
 XX
 OS Homo sapiens.
 XX
 PN US5981700-A.
 XX
 PD 09-NOV-1999.
 XX
 PF 06-JUN-1995; 95US-00468579.
 XX
 PR 17-FEB-1989; 89US-00312543.
 PR 04-DEC-1989; 89US-00441703.
 PR 14-JUN-1991; 91US-00715181.
 PR 08-JUN-1992; 92US-00872646.
 PR 05-MAY-1994; 94US-00239276.
 XX
 PA (FARB) BAYER CORP FORMERLY MOLECULAR DIAGNOSTIC.
 XX
 PI Rabin DU;
 XX
 WPI; 1999-633370/54.
 DR N-PSDB; AAZ32341.
 XX
 PT New pancreatic islet cell antigens, useful in the diagnosis of insulin-
 PT dependent (Type 1) diabetes mellitus.
 XX
 PS Claim 1; Col 83-86; 61pp; English.
 XX
 CC The present sequence represents a human pancreatic islet cell antigen
 CC (ICA) that binds with antibodies found in the sera of patients afflicted
 CC with insulin-dependent (Type 1) diabetes mellitus (IDDM). ICAs from the
 CC present invention are encoded by the DNA insert of a recombinant cloning
 CC vehicle selected from ATCC 40550 (AAZ32333, encoding AAY49847), ATCC
 CC 40553 (AAZ32334, encoding AAY49848), ATCC 40554 (AAZ32335, encoding

CC AAY49849), ATCC 40551 (AAZ32336, encoding AAY49850), ATCC 40552
 CC (AAZ32337, encoding AAY49851), ATCC 40703 (AAZ32338, encoding AAY49852),
 CC ATCC 40704 (AAZ32339, encoding AAY49853), ATCC 40705 (AAZ32340), ATCC
 CC 40706 (AAZ32341, encoding AAY49854) and ATCC 75030 (AAZ32342, encoding
 CC AAY49855). ICA proteins and their peptide fragments can be used in the
 CC diagnosis of IDDM and in detecting or blocking human immunoglobulin, T-
 CC cells or B-cells involved in IDDM
 XX
 SQ Sequence 294 AA;

Query Match 100.0%; Score 24; DB 2; Length 294;
 Best Local Similarity 100.0%; Pred. No. 4e-16;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVKVESSPSRSDYINASPIIEHDP 24
 |||||
 Db 110 KLVKVESSPSRSDYINASPIIEHDP 133

RESULT 3
 AA80486
 ID AAW80486 standard; protein; 294 AA.
 AC AAW80486;
 XX
 DT 01-FEB-1999 (first entry)
 XX
 DE Islet cell antibody antigen encoded by clone ICA-512.
 KW Pancreatic islet cell antibody antigen; ICA antigen; diagnosis;
 KW treatment; insulin dependent diabetes mellitus; IDDM; autoantibody;
 KW autoimmune disease; human.
 XX
 OS Homo sapiens.
 XX
 PN US5840836-A.
 XX
 PD 24-NOV-1998.
 XX
 PF 05-MAY-1994; 94US-00239276.
 XX
 PR 17-FEB-1989; 89US-00312543.
 PR 04-DEC-1989; 89US-00441703.
 PR 14-JUN-1991; 91US-00715181.
 PR 08-JUN-1992; 92US-00872646.
 XX
 PA (FARB) BAYER CORP.
 XX
 PI Rabin DU;
 XX
 WPI; 1999-034118/03.
 DR N-PSDB; AAV63560.
 XX
 PT Polypeptide antigens from pancreatic islet cells - useful for the
 PT detection and treatment of insulin dependent diabetes mellitus.
 XX
 PS Claim 1; Col 51-54; 47pp; English.
 XX
 CC The present sequence represents a pancreatic islet cell antibody (ICA)
 CC antigen. The proteins are used for the diagnosis and treatment of insulin
 CC dependent (Type 1) diabetes mellitus (IDDM). They are recognised by
 CC autoantibodies of the disease as IDDM is now thought to be an autoimmune
 CC disease. They were identified from probing of IDDM serum with human CDNA
 CC libraries expressing protein fragments
 XX
 SQ Sequence 294 AA;

Query Match 100.0%; Score 24; DB 2; Length 294;
 Best Local Similarity 100.0%; Pred. No. 4e-16;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVKVESSPSRSDYINASPIIEHDP 24
 |||||

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2004, 11:13:25 ; Search time 19.0843 Seconds
(without alignments)
83.400 Million cell updates/sec

Title: US-10-783-095-12

Perfect score: 24

Sequence: 1 KLKVESPSRSDYINASPIEHDP 24

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 478139 seqs, 66318000 residues

Word size : 0

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/1/iaa/5A.COMB.pap:*
2: /cgn2_6/prodata/1/iaa/5B.COMB.pap:*
3: /cgn2_6/prodata/1/iaa/6A.COMB.pap:*
4: /cgn2_6/prodata/1/iaa/6B.COMB.pap:*
5: /cgn2_6/prodata/1/iaa/PTUS.COMB.pap:*
6: /cgn2_6/prodata/1/iaa/backfiles1.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	24	100.0	246	3	US-08-884-569A-4
2	24	100.0	294	2	US-08-468-576B-18
3	24	100.0	294	2	US-08-468-579B-18
4	24	100.0	294	3	US-08-468-577B-18
5	24	100.0	548	2	US-08-468-576B-19
6	24	100.0	548	2	US-08-468-579B-19
7	24	100.0	548	3	US-08-468-577B-19
8	24	100.0	979	3	US-08-514-213A-2
9	24	100.0	979	4	US-09-015-399-5
10	22	91.7	22	4	US-09-552-802B-41
11	19	79.2	19	4	US-09-552-802B-41
12	17	70.8	17	4	US-09-552-802B-40
13	12	50.0	12	4	US-09-552-802B-39
14	12	50.0	12	4	US-09-552-802B-38
15	11	45.8	11	4	US-09-552-802B-52
16	9	37.5	246	3	US-08-884-569A-3
17	9	37.5	298	3	US-08-811-481-4
18	9	37.5	298	4	US-08-876-527-4
19	9	37.5	376	3	US-08-811-481-7
20	9	37.5	376	4	US-09-876-527-7
21	9	37.5	641	3	US-08-811-481-2
22	9	37.5	641	4	US-08-876-527-2
23	9	37.5	723	4	US-08-548-159-5
24	9	37.5	818	3	US-08-811-481-22
25	9	37.5	818	4	US-09-876-527-22
26	9	37.5	969	2	US-08-548-159-1
27	9	37.5	986	2	US-08-548-159-3

Sequence 2, Appli
Sequence 16, Appl
Sequence 16, Appl
Sequence 6, Appli
Sequence 6, Appli
Sequence 26, Appl
Sequence 26, Appl
Sequence 24, Appl
Sequence 24, Appl
Sequence 10, Appl
Sequence 10, Appl
Sequence 9, Appli
Sequence 14, Appl
Sequence 9, Appli
Sequence 14, Appl
Sequence 16, Appl
Sequence 16, Appl
Sequence 16, Appl

ALIGNMENTS

RESULT 1

US-08-884-569A-4
; Sequence 4, Application US/08884569A
; Patent No. 6393326
; GENERAL INFORMATION:
; APPLICANT: CHIANG, MING-KO
; APPLICANT: FLANAGAN, JOHN G.
; TITLE OF INVENTION: RECEPTOR TYROSINE PHOSPHATASE, AND USES RELATED THERETO
; FILE REFERENCE: HMV-020.01
; CURRENT APPLICATION NUMBER: US/06/884,569A
; CURRENT FILING DATE: 1997-06-27
; PRIOR APPLICATION NUMBER: 60/021,040
; PRIOR FILING DATE: 1996-07-02
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 246
; TYPE: PRT
; ORGANISM: Mus sp.
US-08-884-569A-4

Query Match 100.0%; Score 24; DB 3; Length 246;
Best Local Similarity 100.0%; Pred. No. 6.6e-17;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLKVESPSRSDYINASPIEHDP 24
Db 19 KLKVESPSRSDYINASPIEHDP 42

RESULT 2

US-08-468-576B-18
; Sequence 18, Application US/08468576B
; Patent No. 5955345
; GENERAL INFORMATION:
; APPLICANT: Rabin, Daniel
; TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
; TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sprung Kramer Schaefer & Briscoe
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: System 7.5

```

; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,576B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/239,276
; FILING DATE: 05-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,646
; FILING DATE: 08-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/715,181
; FILING DATE: 14-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: MDI 251.7-KGB
; FILING DATE: 04-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/312,543
; FILING DATE: 17-FEB-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Kurt G. Briscoe
; REGISTRATION NUMBER: 33,141
; REFERENCE/DOCKET NUMBER: MDI 251.7-KGB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914) 332-1700
; TELEFAX: (914) 332-1844
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 294 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-468-576B-18

Query Match 100.0%; Score 24; DB 2; Length 294;
Best Local Similarity 100.0%; Pred. No. 7.7e-17;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVKVESPSRSDYINASPIIIEHDP 24
Db 110 KLVKVESPSRSDYINASPIIIEHDP 133

RESULT 3
US-08-468-579B-18
; Sequence 18, Application US/08468579B
; Patent No. 5981700
; GENERAL INFORMATION:
; APPLICANT: Rabin, Daniel
; TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
; TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sprung Kramer Schaefer & Briscoe
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: System 7.5
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,579B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/239,276
; FILING DATE: 05-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,646
; FILING DATE: 17-FEB-1989
; APPLICATION NUMBER: US 07/312,543
; FILING DATE: 17-FEB-1989

; US-08-468-579B-18
; Sequence 18, Application US/08468577B
; Patent No. 6001804
; GENERAL INFORMATION:
; APPLICANT: Rabin, Daniel
; TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
; TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sprung Kramer Schaefer & Briscoe
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: System 7.5
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,577B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/239,276
; FILING DATE: 05-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,646
; FILING DATE: 08-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/715,181
; FILING DATE: 14-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/441,703
; FILING DATE: 04-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/312,543
; FILING DATE: 17-FEB-1989

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2004, 11:39:31 ; Search time 47.4217 Seconds
(without alignments)
164.085 Million cell updates/sec

Title: US-10-783-095-12
Perfect score: 24
Sequence: 1 KLVESPSRSDYINASPIIIEHP 24

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1370721 seqs, 324215800 residues

Word size : 0
Total number of hits satisfying chosen parameters: 1370721

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

- Database : Published Applications AA:*
- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
 - 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
 - 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
 - 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
 - 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
 - 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
 - 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
 - 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
 - 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
 - 16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
 - 17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
 - 18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
 - 19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
 - 20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	24	100.0	979	14	US-10-038-686-3
2	22	91.7	22	15	US-10-378-479-41
3	19	79.2	19	15	US-10-378-479-42
4	17	70.8	17	15	US-10-378-479-40
5	12	50.0	12	15	US-10-378-479-38
6	12	50.0	12	15	US-10-378-479-39
7	11	45.8	11	15	US-10-378-479-52
8	9	37.5	298	9	US-09-876-527-4
9	9	37.5	298	14	US-10-124-089-4
10	9	37.5	376	9	US-09-876-527-7
11	9	37.5	376	14	US-10-124-089-7
12	9	37.5	641	9	US-09-876-527-2
13	9	37.5	641	14	US-10-124-089-2

14	9	37.5	818	9	US-09-876-527-22	Sequence 22, Appl
15	9	37.5	818	14	US-10-124-089-22	Sequence 22, Appl
16	9	37.5	1012	9	US-09-876-527-16	Sequence 16, Appl
17	9	37.5	1012	14	US-10-124-089-16	Sequence 16, Appl
18	64	29.2	64	14	US-10-029-386-30320	Sequence 30320, A
19	7	29.2	65	15	US-10-424-599-219467	Sequence 219467, A
20	7	29.2	135	16	US-10-767-701-59696	Sequence 59696, A
21	7	29.2	235	15	US-10-087-684-94	Sequence 94, Appl
22	7	29.2	235	15	US-10-218-779-94	Sequence 94, Appl
23	7	29.2	235	15	US-10-072-012-819	Sequence 819, Appl
24	7	29.2	263	15	US-10-087-684-93	Sequence 93, Appl
25	7	29.2	263	15	US-10-218-779-93	Sequence 93, Appl
26	7	29.2	264	14	US-10-245-539-6	Sequence 6, Appl
27	7	29.2	289	14	US-10-314-232-47	Sequence 47, Appl
28	7	29.2	294	9	US-09-788-626-17	Sequence 27, Appl
29	7	29.2	301	15	US-10-425-114-54172	Sequence 54172, A
30	7	29.2	307	9	US-09-788-626-25	Sequence 25, Appl
31	7	29.2	309	9	US-09-788-626-11	Sequence 11, Appl
32	7	29.2	310	9	US-09-788-626-10	Sequence 10, Appl
33	7	29.2	310	9	US-09-788-626-15	Sequence 15, Appl
34	7	29.2	312	15	US-10-634-027-6	Sequence 6, Appl
35	7	29.2	313	9	US-09-788-626-17	Sequence 17, Appl
36	7	29.2	319	15	US-10-634-027-7	Sequence 7, Appl
37	7	29.2	322	14	US-10-314-232-11	Sequence 11, Appl
38	7	29.2	334	14	US-10-314-232-7	Sequence 7, Appl
39	7	29.2	335	9	US-09-788-626-35	Sequence 35, Appl
40	7	29.2	344	16	US-10-408-765A-1670	Sequence 1670, Ap
41	7	29.2	401	14	US-10-314-233-15	Sequence 15, Appl
42	7	29.2	435	14	US-10-238-075-1224	Sequence 1224, Ap
43	7	29.2	442	9	US-09-925-300-950	Sequence 950, Appl
44	7	29.2	517	15	US-10-425-114-36999	Sequence 36999, A
45	7	29.2				

ALIGNMENTS

RESULT 1
US-10-038-686-3
; Sequence 3, Application US/10038686
; Publication No. US20030045467A1
; GENERAL INFORMATION:
; APPLICANT: Orban, Tihamer
; TITLE OF INVENTION: AUTOANTIGEN VACCINE
; FILE REFERENCE: 10276-067001
; CURRENT APPLICATION NUMBER: US/10/038,686
; CURRENT FILING DATE: 2002-05-31
; PRIOR APPLICATION NUMBER: 60/260,068
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 979
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-038-686-3

Query Match 100.0%; Score 24; DB 14; Length 979;
Best Local Similarity 100.0%; Pred. No. 2.2e-15;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	KLVESPSRSDYINASPIIIEHP	24
Db	752	KLVESPSRSDYINASPIIIEHP	775

RESULT 2
US-10-378-479-41
; Sequence 41, Application US/10378479
; Publication No. US20040006202A1
; GENERAL INFORMATION:
; APPLICANT: Peakman, Mark
; APPLICANT: Chicz, Roman M.

```

; TITLE OF INVENTION: PEPTIDE EPITOPES RECOGNIZED BY DISEASE PROMOTING
; FILE REFERENCE: 08191-009002
; CURRENT APPLICATION NUMBER: US/10/378,479
; CURRENT FILING DATE: 2003-03-03
; PRIOR APPLICATION NUMBER: US/09/552,802B
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US/09/295,868
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: US/09/130,355
; PRIOR FILING DATE: 1999-04-21
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-378-479-41

```

```

Query Match          91.7%; Score 22; DB 15; Length 22;
Best Local Similarity 100.0%; Pred. No. 7.8e-15;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 3 KVESSPSRSDYINASPIIIEHDP 24
   |||||
Db 1 KVESSPSRSDYINASPIIIEHDP 22

```

```

RESULT 3
US-10-378-479-42
; Sequence 42, Application US/10378479
; Publication No. US20040006202A1
; GENERAL INFORMATION:
; APPLICANT: Peakman, Mark
; APPLICANT: Chiciz, Roman M.
; TITLE OF INVENTION: PEPTIDE EPITOPES RECOGNIZED BY DISEASE PROMOTING
; FILE REFERENCE: 08191-009002
; CURRENT APPLICATION NUMBER: US/10/378,479
; CURRENT FILING DATE: 2003-03-03
; PRIOR APPLICATION NUMBER: US/09/552,802B
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US/09/295,868
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: US/09/130,355
; PRIOR FILING DATE: 1999-04-21
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 42
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-378-479-42

```

```

Query Match          79.2%; Score 19; DB 15; Length 19;
Best Local Similarity 100.0%; Pred. No. 6.2e-12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2 LKVESSPSRSDYINASPII 20
   |||||
Db 1 LKVESSPSRSDYINASPII 19

```

```

RESULT 4
US-10-378-479-40
; Sequence 40, Application US/10378479
; Publication No. US20040006202A1
; GENERAL INFORMATION:
; APPLICANT: Peakman, Mark
; APPLICANT: Chiciz, Roman M.
; TITLE OF INVENTION: PEPTIDE EPITOPES RECOGNIZED BY DISEASE PROMOTING
; FILE REFERENCE: 08191-009002

```

```

; CURRENT APPLICATION NUMBER: US/10/378,479
; CURRENT FILING DATE: 2003-03-03
; PRIOR APPLICATION NUMBER: US/09/552,802B
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US/09/295,868
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: US/09/130,355
; PRIOR FILING DATE: 1999-04-21
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 40
; LENGTH: 17
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-378-479-40

```

```

Query Match          70.8%; Score 17; DB 15; Length 17;
Best Local Similarity 100.0%; Pred. No. 5.2e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 KLVESPSRSDYINAS 17
   |||||
Db 1 KLVESPSRSDYINAS 17

```

```

RESULT 5
US-10-378-479-38
; Sequence 38, Application US/10378479
; Publication No. US20040006202A1
; GENERAL INFORMATION:
; APPLICANT: Peakman, Mark
; APPLICANT: Chiciz, Roman M.
; TITLE OF INVENTION: PEPTIDE EPITOPES RECOGNIZED BY DISEASE PROMOTING
; FILE REFERENCE: 08191-009002
; CURRENT APPLICATION NUMBER: US/10/378,479
; CURRENT FILING DATE: 2003-03-03
; PRIOR APPLICATION NUMBER: US/09/552,802B
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US/09/295,868
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: US/09/130,355
; PRIOR FILING DATE: 1999-04-21
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 38
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-378-479-38

```

```

Query Match          50.0%; Score 12; DB 15; Length 12;
Best Local Similarity 100.0%; Pred. No. 3.2e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 3 KVESSPSRSDYI 14
   |||||
Db 1 KVESSPSRSDYI 12

```

```

RESULT 6
US-10-378-479-39
; Sequence 39, Application US/10378479
; Publication No. US20040006202A1
; GENERAL INFORMATION:
; APPLICANT: Peakman, Mark
; APPLICANT: Chiciz, Roman M.
; TITLE OF INVENTION: PEPTIDE EPITOPES RECOGNIZED BY DISEASE PROMOTING
; FILE REFERENCE: 08191-009002
; CURRENT APPLICATION NUMBER: US/10/378,479
; CURRENT FILING DATE: 2003-03-03
; PRIOR APPLICATION NUMBER: US/09/552,802B

```

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 28, 2004, 16:28:53 ; Search time 1574.75 Seconds
(without alignments)
720.721 Million cell updates/sec

Title: US-10-783-095-12

Perfect score: 24

Sequence: 1 KLVKVESSPSRSDYINASPIIBHP 24

Scoring table:

OLIGO
Xgapop 60.0, Xgapext 60.0
Ygapop 60.0, Ygapext 60.0
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4526729 seqs, 23644849745 residues

Word size: 1

Total number of hits satisfying chosen parameters: 9043826

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-Q=/cgn2.1/USPto.spool.p/US10783095/runat.26102004.185454.25032/app.query.fasta_1.796
-DS=GenEmbl -QFMT=fasta5 -SUFFIX=oli.rge -MINMATCH=0.1 -LOPCL=0 -LOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10783095 @cgn1.1.6148 @runat.26102004.185454.25032 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -XGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl:

1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_ses:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	24	100.0	1210	6	AR369579 Sequence
2	24	100.0	1210	6	AR404158 Sequence
3	24	100.0	1413	6	AR060650 Sequence
4	24	100.0	1413	6	AR085261 Sequence

5	24	100.0	1413	6	AR094588	Sequence
6	24	100.0	1635	10	BC064020	Mus muscu
7	24	100.0	1776	9	BT006975	Homo sapi
8	24	100.0	1776	12	BT007693	Synthetic
9	24	100.0	2296	9	AK122679	Homo sapi
10	24	100.0	2321	10	RA12683	Rat mRNA fo
11	24	100.0	2674	10	AF332084	Mus muscu
12	24	100.0	2683	10	AF332083	Mus muscu
13	24	100.0	2949	4	AF075170	Bos tauru
14	24	100.0	2952	10	RN10105	R. norvegicu
15	24	100.0	3037	6	CQ114777	Sequence
16	24	100.0	3056	9	BC007713	Homo sapi
17	24	100.0	3311	6	AR060651	Sequence
18	24	100.0	3311	6	AR085262	Sequence
19	24	100.0	3311	9	AR094589	Sequence
20	24	100.0	3311	9	HS1CA512	Homo sapien
21	24	100.0	3396	6	A37107	Sequence 1
22	24	100.0	3397	10	RN040652	Rattus norv
23	24	100.0	3446	6	CQ812832	Sequence
24	24	100.0	3446	10	MMU11812	Mus musculu
25	24	100.0	3477	10	RATPDPTPLP	Rattus sp.
26	24	100.0	3485	10	BC068165	Mus muscu
27	24	100.0	3542	6	AX364925	Sequence
28	24	100.0	3556	9	BC070053	Homo sapi
29	24	100.0	3561	6	A37109	Sequence 3
30	24	100.0	3566	6	AX305306	Sequence
31	24	100.0	3566	10	MMTP35A	X74438 M.musculus
32	24	100.0	3613	9	HUMTYRPHO	Homo sapien
33	24	100.0	3637	6	AR454532	Sequence
34	19	79.2	590	11	EV106091	MARC 8095
35	19	79.2	61579	9	AC114803	Homo sapi
36	19	79.2	163300	2	AC060820	Homo sapi
37	19	79.2	189793	2	AC149188	Papio anu
38	19	79.2	220494	2	AC121001	Rattus no
39	19	79.2	221015	2	AC106467	Rattus no
40	19	79.2	244567	2	AC118950	Rattus no
41	11	45.8	190715	5	EX247878	Zebrafish
42	11	45.8	199832	2	EX232576	Danio rer
43	11	45.8	237129	5	EX005327	Zebrafish
44	9	37.5	201	11	BV203767	sgnm21297
45	9	37.5	758	6	AR369581	Sequence

ALIGNMENTS

RESULT 1
AR369579
LOCUS AR369579 1210 bp mRNA linear PAT 12-SEP-2003
DEFINITION Sequence 30 from patent US 6300093.
ACCESSION AR369579
VERSION AR369579.1 GI:34605766
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.

REFERENCE
1 (bases 1 to 1210)
AUTHORS Kindvogel,W., Jelinek,L.J., Sheppard,P.O., Hagopian,W.A. and
Iacasse,J.M.
TITLE Islet cell antigen 1851
JOURNAL Patent: US 6300093-A 30 09-OCT-2001;
FEATURES Location/Qualifiers
source 1..1210
/organism="unknown"
/mol_type="mRNA"

ORIGIN

Alignment Scores:
Pred. No.: 1.21e-14 Length: 1210
Score: 24.00 Matches: 24
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-783-095-12 (1-24) x AR369579 (1-1210)

QY 1 LysLeuLysValGluSerSerProSerArgSerAspTyrIleAsnAlaSerProIleile 20
 DB 527 AAACCTGAAGGTGGAGAGCAGCCCTTCTCGAGCGATTACATCAACGCCAGCCCATTTATT 586

QY 21 GluHisAspPro 24

DB 587 GAGCATGACCCCT 598

RESULT 2
 AR404158

LOCUS AR404158 1210 bp mRNA linear PAT 18-DEC-2003

DEFINITION Sequence 30 from patent US 6627735.

ACCESSION AR404158

VERSION AR404158.1 GI:40152191

KEYWORDS

SOURCE

ORGANISM

REFERENCE 1 (bases 1 to 1210)

AUTHORS Kindsvogel, W., Jelinek, L.J., Sheppard, P.O., Hagopian, W.A. and

Lagasse, J.M.

TITLE Islet cell antigen 1851

JOURNAL Patent: US 6627735-A 30 30-SEP-2003;

FEATURES

1..1210

/organism="unknown"

/mol_type="rRNA"

ORIGIN

Alignment Scores:

Pred. No.: 1.21e-14 Length: 1210

Score: 24.00 Matches: 24

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 6 Gaps: 0

US-10-783-095-12 (1-24) x AR404158 (1-1210)

QY 1 LysLeuLysValGluSerSerProSerArgSerAspTyrIleAsnAlaSerProIleile 20
 DB 527 AAACCTGAAGGTGGAGAGCAGCCCTTCTCGAGCGATTACATCAACGCCAGCCCATTTATT 586

QY 21 GluHisAspPro 24

DB 587 GAGCATGACCCCT 598

RESULT 3
 AR060650

LOCUS AR060650 1413 bp DNA linear PAT 29-SEP-1999

DEFINITION Sequence 9 from patent US 5840836.

ACCESSION AR060650

VERSION AR060650.1 GI:5987100

KEYWORDS

SOURCE

ORGANISM

REFERENCE 1 (bases 1 to 1413)

AUTHORS Rabin, D.U.

TITLE Pancreatic islet cell antigens obtained by molecular cloning

JOURNAL Patent: US 5840836-A 9 24-NOV-1998;

FEATURES

1..1413

/organism="unknown"

/mol_type="unassigned DNA"

ORIGIN

Alignment Scores:

Pred. No.: 1.39e-14 Length: 1413

Score: 24.00 Matches: 24

Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-10-783-095-12 (1-24) x AR060650 (1-1413)

QY 1 LysLeuLysValGluSerSerProSerArgSerAspTyrIleAsnAlaSerProIleile 20
 DB 330 AAACCTGAAGGTGGAGAGCAGCCCTTCTCGAGCGATTACATCAACGCCAGCCCATTTATT 389

QY 21 GluHisAspPro 24

DB 390 GAGCATGACCCCT 401

RESULT 4
 AR085261

LOCUS AR085261 1413 bp DNA linear PAT 01-SEP-2000

DEFINITION Sequence 9 from patent US 5981700.

ACCESSION AR085261

VERSION AR085261.1 GI:10012031

KEYWORDS

SOURCE

ORGANISM

REFERENCE 1 (bases 1 to 1413)

AUTHORS Rabin, D.U.

TITLE Pancreatic islet cell antigens obtained by molecular cloning

JOURNAL Patent: US 5981700-A 9 09-NOV-1999;

FEATURES

1..1413

/organism="unknown"

/mol_type="unassigned DNA"

ORIGIN

Alignment Scores:

Pred. No.: 1.39e-14 Length: 1413

Score: 24.00 Matches: 24

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 6 Gaps: 0

US-10-783-095-12 (1-24) x AR085261 (1-1413)

QY 1 LysLeuLysValGluSerSerProSerArgSerAspTyrIleAsnAlaSerProIleile 20
 DB 330 AAACCTGAAGGTGGAGAGCAGCCCTTCTCGAGCGATTACATCAACGCCAGCCCATTTATT 389

QY 21 GluHisAspPro 24

DB 390 GAGCATGACCCCT 401

RESULT 5
 AR094588

LOCUS AR094588 1413 bp DNA linear PAT 08-SEP-2000

DEFINITION Sequence 9 from patent US 6001804.

ACCESSION AR094588

VERSION AR094588.1 GI:10021642

KEYWORDS

SOURCE

ORGANISM

REFERENCE 1 (bases 1 to 1413)

AUTHORS Rabin, D.U.

TITLE Pancreatic islet cell antigens obtained by molecular cloning

JOURNAL Patent: US 6001804-A 9 14-DEC-1999;

FEATURES

1..1413

/organism="unknown"

/mol_type="unassigned DNA"

ORIGIN

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 28, 2004, 16:24:37 ; Search time 312.578 Seconds
(without alignments)
403.055 Million cell updates/sec

Title: US-10-783-095-12

Perfect score: 24

Sequence: 1 KLVKVESSPSRSDYINASPIIEHDP 24

Scoring table:
Xgapop 60.0, Xgapext 60.0
Ygapop 60.0, Ygapext 60.0
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4134886 seqs, 2624710521 residues

Word size: 1

Total number of hits satisfying chosen parameters: 8264151

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xlp
-Q/cgn2.1/US10783095/runat 26102004 185454 25024/app query.fasta_1.796
-DB=N_Geneseq 23Sep04 -OPT=fastap -SUFFIX=oli.rng -MINMATCH=0.1 -LOOPT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALLCN=200 -THR SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pro -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10783095 @CNC 1.1 1241 @runat 26102004 185454 25024 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N_Geneseq 23Sep04:*

- 1: Geneseqn1980s:*
- 2: Geneseqn1990s:*
- 3: Geneseqn2000s:*
- 4: Geneseqn2001as:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002as:*
- 7: Geneseqn2002bs:*
- 8: Geneseqn2003as:*
- 9: Geneseqn2003bs:*
- 10: Geneseqn2003cs:*
- 11: Geneseqn2003ds:*
- 12: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	24	100.0	1210	2 AAT95246	Aat95246 Human IA-
2	24	100.0	1413	2 AA209922	Aa209922 Human isl
3	24	100.0	1413	2 AA232341	Aa232341 Human pan
4	24	100.0	1413	2 AAV63560	Aav63560 Islet cel
5	24	100.0	1413	3 AA46459	Aa46459 Pancreat
6	24	100.0	3311	2 AA209923	Aa209923 Human isl

7	24	100.0	3311	2 AA232342	Aa232342 Human pan
8	24	100.0	3311	2 AAV63561	Aav63561 Islet cel
9	24	100.0	3311	3 AA246460	Aa246460 Pancreat
10	24	100.0	3396	2 AAQ55515	Aa25515 PTPase PT
11	24	100.0	3446	12 ADO60053	Ado60053 CRH signa
12	24	100.0	3477	10 ABT41712	Abt41712 Toxicity
13	24	100.0	3542	6 ABL39766	Ab139766 Human NS
14	24	100.0	3561	2 AAQ55970	Aa255970 PTPase PT
15	24	100.0	3566	6 ABI99237	Abi99237 Mouse isc
16	24	100.0	3613	2 AAX87627	Aax87627 Human isl
17	24	100.0	3613	4 AAI68754	Aai68754 Human aut
18	24	100.0	4685	4 AAI68774	Aai68774 Fusion pr
19	24	100.0	5920	9 AAI68773	Aai68773 Fusion pr
20	10	41.7	453	9 ACH14280	Ach14280 Human adu
21	9	37.5	758	2 AAT95226	Aat95226 Human isl
22	9	37.5	894	2 AAT95223	Aat95223 Human isl
23	9	37.5	1110	2 AAT67297	Aat67297 Type I di
24	9	37.5	1150	2 AAT95228	Aat95228 Human isl
25	9	37.5	1183	2 AAT95229	Aat95229 Human isl
26	9	37.5	2171	2 AAT95222	Aat95222 Macaque i
27	9	37.5	2351	2 AAT88480	Aat88480 Mouse ins
28	9	37.5	2464	2 AAT95221	Aat95221 Human isl
29	9	37.5	2736	2 AAT95231	Aat95231 Macaque i
30	9	37.5	2840	2 AAT67295	Aat67295 Type I di
31	9	37.5	2943	2 AAT88478	Aat88478 Human ins
32	9	37.5	2994	2 AAT88479	Aat88479 Human ins
33	9	37.5	3200	2 AAV04936	Aav04936 Mouse pro
34	9	37.5	3287	2 AAT95220	Aat95220 Macaque i
35	9	37.5	3310	2 AAT67296	Aat67296 Type I di
36	9	37.5	3856	5 AAS87632	Aas87632 DNA encod
37	9	37.5	4136	2 AAT95368	Aat95368 Human pro
38	9	37.5	4699	10 ADD14680	Add14680 Human SRC
39	9	37.5	4719	12 ADP18660	Adp18660 Human TAR
40	9	37.5	4767	11 ADN95437	Adn95437 Human BEC
41	9	37.5	4856	5 AAS87633	Aas87633 DNA encod
42	8	33.3	179	10 ACA55641	Acas55641 Human sig
43	8	33.3	179	12 ADI55437	Adi55437 Human pol
C 44	8	33.3	986	6 AAL44150	Aal44150 Barley ye
C 45	8	33.3	2724	3 AA288566	Aaz88566 B2DV-PAV

ALIGNMENTS

RESULT 1

AAT95246
ID AAT95246 standard; cDNA; 1210 BP.

XX AC AAT95246;

XX DT 27-MAR-1998 (first entry)

XX DE Human IA-2/ICAS12 protein tyrosine phosphatase.

XX KW Islet cell antigen 185i; autoantigen; protein tyrosine phosphatase;
KW insulin-dependent diabetes mellitus; IDDM; diagnosis; therapy; human;
KW IA-2/ICAS12; ds.

XX OS Homo sapiens.

XX FN W09732984-A1.

XX PD 12-SEP-1997.

XX PF 05-MAR-1997; 97WO-US0003532.

XX PR 06-MAR-1996; 96US-0012927P.

XX PR 15-OCT-1996; 96US-0027540P.

XX PA (ZYMO) ZYMOGENETICS INC.

XX PA (UNIW) UNIV WASHINGTON.

XX PI Kindsvogel W, Jelinek L, Sheppard PO, Hagopian W, Lagasse J;

DR WPI; 1997-457535/42.
 XX
 PT Mammalian pancreatic islet cell antigen and related DNA - used to detect
 PT auto:antibodies indicative of insulin-dependent diabetes or pre-
 PT disposition to it.
 XX
 PS Example 3; Page 111; 134pp; English.
 XX
 CC This cDNA sequence comprises the cytoplasmic region of human protein
 CC tyrosine phosphatase IA-2/ICA512. It was obtained from U87MG glioblastoma
 CC cells (ATCC M85) by PCR amplification (see AAT95244-45). The 1.2 kb
 CC product was ligated into pCEP. Sera from prediabetic subjects were
 CC screened for IA-2/ICA512 autoantibodies. A novel islet cell antigen,
 CC designated 1851 (see AAT95296-97), has been identified that forms an
 CC immune complex with an autoantibody found in patients at risk of, or
 CC predisposed to, insulin-dependent diabetes mellitus
 XX
 SQ Sequence 1210 BP; 245 A; 383 C; 370 G; 212 T; 0 U; 0 Other;
 XX
 Alignment Scores:
 Pred. No.: 4,858-15 Length: 1210
 Score: 24.00 Matches: 24
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 2 Gaps: 0
 XX
 US-10-783-095-12 (1-24) x AAT95246 (1-1210)
 QY 1 LysLeuLysValGluSerProSerArgSerAspTyrIleAsnAlaSerProIleile 20
 Db 527 AAACCTGAAGTGCAGAGAGCCCTTCTCGAGCGATTACATCAACGCCAGCCCATTTATT 586
 QY 21 GluHisAspPro 24
 Db 587 GAGCATGACCCCT 598
 XX
 RESULT 2
 ID AAZ09922
 AC AAZ09922 standard; cDNA; 1413 BP.
 AC AAZ09922;
 DT 29-NOV-1999 (first entry)
 DE Human islet cell antigen clone ICA-512 cDNA.
 KW Islet cell antigen; ICA; human; pancreas; antibody; IDDM; immunoassay;
 KW insulin-dependent type I diabetes mellitus; diagnosis; immunoglobulin;
 KW T-cell; B-cell; islet cell destruction; immunoreactivity; disease;
 KW immune therapy; stimulating antigen; screening; ds.
 XX
 OS Homo sapiens.
 FH Location/Qualifiers
 FT 1..886
 FT /*tag= a
 FT /codon_start= 3
 FT /note= "Partial sequence, no start codon given"
 FT /product= "ICA-512"
 XX
 US9595345-A.
 XX
 PD 21-SEP-1999.
 XX
 PF 06-JUN-1995; 95US-00468576.
 PR 17-FEB-1989; 89US-00312543.
 PR 04-DEC-1989; 89US-00441703.
 PR 14-JUN-1991; 91US-00715181.
 PR 08-JUN-1992; 92US-00872646.
 PR 05-MAY-1994; 94US-00239276.
 XX

PA (FARB) BAYER CORP FORMERLY MOLECULAR DIAGNOSTIC.
 XX
 PI Rabin DU;
 XX
 DR WPI; 1999-560494/47.
 DR P-PSDB; AAY33360.
 XX
 PT Isolated DNA encoding antigens reactive with serum from diabetics useful
 PT to develop products for the diagnosis and monitoring of insulin-dependent
 PT diabetes mellitus.
 XX
 PS Claim 1; Col 51-56; 63pp; English.
 XX
 CC This invention describes novel isolated and purified DNAs encoding
 CC pancreatic islet cell antigens reactive with islet cell antibodies found
 CC in sera of patients with insulin-dependent type I diabetes mellitus
 CC (IDDM). The polypeptides of the invention are useful as immunoassay
 CC reagents in the presymptomatic diagnosis of insulin-dependent type I
 CC diabetes mellitus (IDDM). The ability of the polypeptides to bind the
 CC antibody binding site on islet cell antibodies (ICAs) also confers
 CC utility in the binding or blocking of human immunoglobulin, T-cells, or B
 CC -cells involved in IDDM. The products can also be used to study the
 CC biological mechanisms involved in islet cell destruction and the
 CC appearance of ICA. The immunoreactivity profile with different antigens
 CC can provide diagnostically significant information concerning the nature
 CC of the disease, e.g. subtypes, the state of the disease, the proximity to
 CC onset of the disease, and the efficacy of therapy, e.g. immune therapy.
 CC The antigens can serve as stimulating antigens for T-cell culture, and
 CC permitting significantly improved T-cell cloning, identification, and
 CC growth. The availability of large quantities of pure antigen enables the
 CC development of highly sensitive and specific immunoassays which can be
 CC used to screen the general population of presymptomatic IDDM or a
 CC predisposition to develop IDDM. This sequence encodes the human islet
 CC cell antigen from clone ICA-512
 XX
 SQ Sequence 1413 BP; 289 A; 459 C; 412 G; 253 T; 0 U; 0 Other;
 XX
 Alignment Scores:
 Pred. No.: 5,618-15 Length: 1413
 Score: 24.00 Matches: 24
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 2 Gaps: 0
 XX
 US-10-783-095-12 (1-24) x AAZ09922 (1-1413)
 QY 1 LysLeuLysValGluSerProSerArgSerAspTyrIleAsnAlaSerProIleile 20
 Db 330 AAACCTGAAGTGCAGAGAGCCCTTCTCGAGCGATTACATCAACGCCAGCCCATTTATT 389
 QY 21 GluHisAspPro 24
 Db 390 GAGCATGACCCCT 401
 XX
 RESULT 3
 ID AAZ32341
 ID AAZ32341 standard; cDNA; 1413 BP.
 AC AAZ32341;
 XX
 XX 21-JAN-2000 (first entry)
 XX
 DE Human pancreatic islet cell antigen ICA-512 ATCC 40706 cDNA sequence.
 KW Human; pancreatic islet cell antigen; ICA clone; IDDM; diabetes;
 KW insulin dependent type I diabetes mellitus; diagnosis; detection;
 KW immunoglobulin; T-cell; B-cell; antibody binding; ss.
 OS Homo sapiens.
 PN US5981700-A.
 XX

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 28, 2004, 20:02:28 ; Search time 71.1325 Seconds
(without alignments)
239.819 Million cell updates/sec

Title: US-10-783-095-12
Perfect score: 24
Sequence: 1 KLVESPSRSDYINASPIIHDP 24

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 824507 seqs, 355394441 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1642088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters: -DB=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10783095/runat_26102004.185455.25065/app_query.fasta_1.796
-DB=issued_Patents_NA -QMT=fastap -SUFFIX=oli.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=BITS -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdd
-LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10783095 @CEN 1 1.213 @runat 26102004.185455.25065 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -Fgapop=6
-Fgapext=7 -Ygapop=60 -Ygapext=60 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:
5: /cgn2_6/ptodata/1/ina/FACTUS.COMB.seq:
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	24	100.0	1210	3	US-08-811-481-30 Sequence 30, Appl
2	24	100.0	1210	4	US-09-876-527-30 Sequence 30, Appl
3	24	100.0	1413	2	US-08-239-276-9 Sequence 9, Appl
4	24	100.0	1413	2	US-08-468-5798-9 Sequence 9, Appl
5	24	100.0	1413	3	US-08-468-5778-9 Sequence 9, Appl
6	24	100.0	3311	2	US-08-239-276-10 Sequence 10, Appl
7	24	100.0	3311	2	US-08-468-5798-10 Sequence 10, Appl
8	24	100.0	3311	3	US-08-468-5778-10 Sequence 10, Appl
9	24	100.0	3613	3	US-08-514-213A-1 Sequence 1, Appl
10	24	100.0	3613	4	US-09-015-399-9 Sequence 9, Appl
11	24	100.0	3637	4	US-09-566-921-5 Sequence 5, Appl
12	9	37.5	758	3	US-08-811-481-32 Sequence 32, Appl

Sequence 32, Appl
Sequence 3, Appl
Sequence 3, Appl
Sequence 33, Appl
Sequence 33, Appl
Sequence 6, Appl
Sequence 6, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 6, Appl
Sequence 21, Appl
Sequence 21, Appl
Sequence 23, Appl
Sequence 23, Appl
Sequence 2, Appl
Sequence 4, Appl
Sequence 1, Appl
Sequence 15, Appl
Sequence 15, Appl
Sequence 239, Appl
Sequence 2, Appl
Sequence 6, Appl
Sequence 13160, A
Sequence 11236, A
Sequence 1375, Ap
Sequence 1375, Ap
Sequence 1375, Ap
Sequence 1375, Ap
Sequence 32217, A
Sequence 10074, A
Sequence 11566, A
Sequence 14755, A

ALIGNMENTS

RESULT 1
US-08-811-481-30
; Sequence 30, Application US/08811481
; Patent No. 6300093
; GENERAL INFORMATION:
; APPLICANT: Kindsvogel, Wayne
; APPLICANT: Jelinek, Laura J.
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Hagopian, William A.
; APPLICANT: LaGasse, James M.
; TITLE OF INVENTION: ISLET CELL ANTIGEN 1851
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,481
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Lindenfelder, Susan
; REGISTRATION NUMBER: P-41,156
; REFERENCE/DOCKET NUMBER: 95-36
; TELECOMMUNICATION INFORMATION:

```

; TELEPHONE: 206-442-6675
; TELEFAX: 206-442-6678
; TELEX:
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1210 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-811-481-30

Alignment Scores:
Pred. No.: 1,52e-15 Length: 1210
Score: 24.00 Matches: 24
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-783-095-12 (1-24) x US-08-811-481-30 (1-1210)
QY 1 LysLeuLysValGluSerSerProSerArgSerAspTyrIleAsnAlaSerProIleIle 20
Db 527 AAACCTGAAGGTGGAGAGCGCCCTTCGAGCGGATTACATCAACGCCGCCCATTTATT 586

QY 21 GluHisAspPro 24
Db 587 GAGCATGACCCT 598

RESULT 2
US-08-876-527-30
; Sequence 30, Application US/09876527
; Patent No. 6627735
; GENERAL INFORMATION:
; APPLICANT: Kindvogel, Wayne
; Jelinek, Laura J.
; Sheppard, Paul O.
; Hagopian, William A.
; LaGasse, James M.
; TITLE OF INVENTION: ISLET CELL ANTIGEN 1851
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/876,527
; FILING DATE: 07-Jun-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/811,481
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Lingenfelter, Susan
; REGISTRATION NUMBER: P-41,156
; REFERENCE/DOCKET NUMBER: 95-36
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6675
; TELEFAX: 206-442-6678
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1210 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
;
; TELEPHONE: 206-442-6675
; TELEFAX: 206-442-6678
; TELEX:
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1210 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
;
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 30:
US-09-876-527-30

Alignment Scores:
Pred. No.: 1,52e-15 Length: 1210
Score: 24.00 Matches: 24
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-10-783-095-12 (1-24) x US-09-876-527-30 (1-1210)
QY 1 LysLeuLysValGluSerSerProSerArgSerAspTyrIleAsnAlaSerProIleIle 20
Db 527 AAACCTGAAGGTGGAGAGCGCCCTTCGAGCGGATTACATCAACGCCGCCCATTTATT 586

QY 21 GluHisAspPro 24
Db 587 GAGCATGACCCT 598

RESULT 3
US-08-239-276-9
; Sequence 9, Application US/08239276
; Patent No. 5840836
; GENERAL INFORMATION:
; APPLICANT: Rabin, Daniel
; TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sprung Kramer Schaefer & Briscoe
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: System 7.5
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/239,276
; FILING DATE: 05-MAY-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,646
; FILING DATE: 08-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/715,181
; FILING DATE: 14-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/441,703
; FILING DATE: 04-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/312,543
; FILING DATE: 17-FEB-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Kurt G. Briscoe
; REGISTRATION NUMBER: 33,141
; REFERENCE/DOCKET NUMBER: MDI 251.4-KGB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914) 332-1700
; TELEFAX: (914) 332-1844
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1413 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single

```


GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 28, 2004, 22:07:29 ; Search time 1265.93 Seconds.
(without alignments)
97.211 Million cell updates/sec

Title: US-10-783-095-12

Perfect score: 24
Sequence: 1 KLVKVESPSRSDYINASPIEHDP 24

Scoring table:

Gapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3413475 seqs, 2563800928 residues

Word size: 1

Total number of hits satisfying chosen parameters: 6815939

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+pn2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_P/US10783095/runat_26102004_185456_25148/app_query.fasta_1.796
-DB=Published Applications NA -QMT=fastap -SUFFIX=oli.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNIGS=bits -START=1 -END=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEARSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10783095 -CGN_1_1189 -runat_26102004_185456_25148
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=120 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/US07_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US10F_NEW_PUB.seq.*
19: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
20: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
21: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

*

Result No.	Score	Query Match	Length	DB ID	Description
1	24	100.0	1210	9	US-09-876-527-30
2	24	100.0	1210	15	US-10-124-089-30
3	24	100.0	3397	14	US-10-175-523-167
4	24	100.0	3477	16	US-10-152-319A-1414
5	10	41.7	453	10	US-09-918-995-1492
6	9	37.5	758	9	US-09-876-527-32
7	9	37.5	758	15	US-10-124-089-32
8	9	37.5	894	9	US-09-876-527-3
9	9	37.5	894	15	US-10-124-089-3
10	9	37.5	1150	9	US-09-876-527-33
11	9	37.5	1150	15	US-10-124-089-33
12	9	37.5	1163	9	US-09-876-527-6
13	9	37.5	1163	15	US-10-124-089-6
14	9	37.5	2171	9	US-09-876-527-1
15	9	37.5	2171	15	US-10-124-089-1
16	9	37.5	2464	9	US-09-876-527-21
17	9	37.5	2464	15	US-10-124-089-21
18	9	37.5	2736	9	US-09-876-527-23
19	9	37.5	2736	15	US-10-124-089-23
20	9	37.5	3287	9	US-09-876-527-15
21	9	37.5	3287	15	US-10-124-089-15
22	8	33.3	179	16	US-10-305-720-239
23	8	33.3	392	16	US-10-242-53A-34437
24	8	33.3	392	16	US-10-085-78A-34437
25	8	33.3	463	16	US-10-424-599-3767
26	8	33.3	464	18	US-10-674-124A-24805
27	8	33.3	605	13	US-10-027-632-280404
28	8	33.3	605	15	US-10-027-632-280404
29	8	33.3	752	13	US-10-027-632-148982
30	8	33.3	752	15	US-10-027-632-148982
31	8	33.3	878	16	US-10-425-114-17359
32	8	33.3	979	16	US-10-425-114-17358
33	8	33.3	986	9	US-09-984-863-2
34	8	33.3	68495	17	US-10-322-281-750
35	8	33.3	2731748	17	US-10-297-455A-1
36	7	29.2	25	14	US-10-215-112-5302
37	7	29.2	97	16	US-10-210-550-94
38	7	29.2	195	15	US-10-029-386-17615
39	7	29.2	206	9	US-09-783-590-8672
40	7	29.2	206	17	US-10-437-963-75067
41	7	29.2	220	16	US-10-424-599-76625
42	7	29.2	295	9	US-09-736-457-1375
43	7	29.2	295	9	US-09-302-941-1375
44	7	29.2	295	9	US-09-849-626-1375
45	7	29.2	295	14	US-10-017-754-1375

ALIGNMENTS

RESULT 1

US-09-876-527-30
; Sequence 30, Application US/09876527
; Patent No. US20020102616A1
; GENERAL INFORMATION:
; APPLICANT: Kindvogel, Wayne
; Jelinek, Laura J.
; Sheppard, Paul O.
; Hagopian, William A.
; LaGasse, James M.
; TITLE OF INVENTION: ISLET CELL ANTIGEN 1851
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette

```

/ / COMPUTER: IBM Compatible
/ / OPERATING SYSTEM: DOS
/ / SOFTWARE: FastSeq for Windows Version 2.0
/ / CURRENT APPLICATION DATA:
/ / APPLICATION NUMBER: US/09/876,527
/ / FILING DATE: 07-Jun-2001
/ / CLASSIFICATION: <Unknown>
/ / PRIOR APPLICATION DATA:
/ / APPLICATION NUMBER: 08/811,481
/ / FILING DATE: <Unknown>
/ / ATTORNEY/AGENT INFORMATION:
/ / NAME: Lingenfelter, Susan
/ / REGISTRATION NUMBER: P-41,156
/ / REFERENCE/DOCKET NUMBER: 95-36
/ / TELECOMMUNICATION INFORMATION:
/ / TELEPHONE: 206-442-6675
/ / TELEFAX: 206-442-6678
/ / TELEX: <Unknown>
/ / INFORMATION FOR SEQ ID NO: 30:
/ / SEQUENCE CHARACTERISTICS:
/ / LENGTH: 1210 base pairs
/ / TYPE: nucleic acid
/ / STRANDEDNESS: double
/ / TOPOLOGY: linear
/ / MOLECULE TYPE: cDNA
/ / SEQUENCE DESCRIPTION: SEQ ID NO: 30:
US-09-876-527-30

Alignment Scores:
Pred. No.: 6.7e-16 Length: 1210
Score: 24.00 Matches: 24
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-10-783-095-12 (1-24) x US-09-876-527-30 (1-1210)

Qy 1 LysLeuLysValGluSerProSerArgSerAspTyrIleAsnAlaSerProIleIle 20
Db 527 AAACGTGAGGTGGAGGAGGAGCCCTTCTCGAGCGATTACATCAAGCGCCCATTTATT 586

Qy 21 GluHisAspPro 24
Db 587 GAGCATGACCCCT 598

RESULT 2
US-10-124-089-30
/ / Sequence 30, Application US/10124089
/ / Publication No. US2003016067A1
/ / GENERAL INFORMATION:
/ / APPLICANT: Kindsvogel, Wayne
/ / Jelinek, Laura J.
/ / Sheppard, Paul O.
/ / Hagopian, William A.
/ / Lacasse, James M.
/ / TITLE OF INVENTION: ISLET CELL ANTIGEN 1851
/ / NUMBER OF SEQUENCES: 34
/ / CORRESPONDENCE ADDRESS:
/ / ADDRESSEE: ZymoGenetics, Inc.
/ / STREET: 1201 Eastlake Avenue East
/ / CITY: Seattle
/ / STATE: WA
/ / COUNTRY: USA
/ / ZIP: 98102
/ / COMPUTER READABLE FORM:
/ / MEDIUM TYPE: Diskette
/ / COMPUTER: IBM Compatible
/ / OPERATING SYSTEM: DOS
/ / SOFTWARE: FastSeq for Windows Version 2.0
/ / APPLICATION DATA:
/ / APPLICATION NUMBER: US/10/124,089
/ / FILING DATE: 16-Apr-2002

```

```

/ / CLASSIFICATION: <Unknown>
/ / PRIOR APPLICATION DATA:
/ / APPLICATION NUMBER: US/08/811,481
/ / FILING DATE: <Unknown>
/ / ATTORNEY/AGENT INFORMATION:
/ / NAME: Lingenfelter, Susan
/ / REGISTRATION NUMBER: P-41,156
/ / REFERENCE/DOCKET NUMBER: 95-36
/ / TELECOMMUNICATION INFORMATION:
/ / TELEPHONE: 206-442-6675
/ / TELEFAX: 206-442-6678
/ / TELEX: <Unknown>
/ / INFORMATION FOR SEQ ID NO: 30:
/ / SEQUENCE CHARACTERISTICS:
/ / LENGTH: 1210 base pairs
/ / TYPE: nucleic acid
/ / STRANDEDNESS: double
/ / TOPOLOGY: linear
/ / MOLECULE TYPE: cDNA
/ / SEQUENCE DESCRIPTION: SEQ ID NO: 30:
US-10-124-089-30

Alignment Scores:
Pred. No.: 6.7e-16 Length: 1210
Score: 24.00 Matches: 24
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 15 Gaps: 0

US-10-783-095-12 (1-24) x US-10-124-089-30 (1-1210)

Qy 1 LysLeuLysValGluSerProSerArgSerAspTyrIleAsnAlaSerProIleIle 20
Db 527 AAACGTGAGGTGGAGGAGGAGCCCTTCTCGAGCGATTACATCAAGCGCCCATTTATT 586

Qy 21 GluHisAspPro 24
Db 587 GAGCATGACCCCT 598

RESULT 3
US-10-175-523-167
/ / Sequence 167, Application US/10175523
/ / Publication No. US20030096264A1
/ / GENERAL INFORMATION:
/ / APPLICANT: Brockman, Jeffrey
/ / APPLICANT: Evans, David
/ / APPLICANT: Hook, Derek
/ / APPLICANT: Klimczak, Leszek
/ / APPLICANT: Laerg, Pascal
/ / APPLICANT: Palfreyman, Michael
/ / APPLICANT: Rajan, Prithi
/ / TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)
/ / FILE REFERENCE: 3235/1J795-US3
/ / CURRENT APPLICATION NUMBER: US/10/175,523
/ / CURRENT FILING DATE: 2002-06-18
/ / PRIOR APPLICATION NUMBER: US 60/299,151
/ / PRIOR FILING DATE: 2001-06-18
/ / PRIOR APPLICATION NUMBER: US 60/317,828
/ / PRIOR FILING DATE: 2001-09-07
/ / PRIOR APPLICATION NUMBER: US 60/325,150
/ / PRIOR FILING DATE: 2001-09-25
/ / PRIOR APPLICATION NUMBER: US 60/333,047
/ / PRIOR FILING DATE: 2001-11-14
/ / PRIOR APPLICATION NUMBER: US 60/349,936
/ / PRIOR FILING DATE: 2002-01-18
/ / PRIOR APPLICATION NUMBER: US 60/361,834
/ / PRIOR FILING DATE: 2002-03-04
/ / NUMBER OF SEQ ID NOS: 197
/ / SOFTWARE: Patent in version 3.1
/ / SEQ ID NO 167
/ / LENGTH: 3397
/ / TYPE: DNA

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2004, 11:13:10 ; Search time 12.4337 Seconds
(without alignments)
185.721 Million cell updates/sec

Title: US-10-783-095-12

Perfect score: 24

Sequence: 1 KLKVSSPSRSDYINASPIIEHDP 24

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283416 seqs, 96216763 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

PIR 79:*

1: Pirl:*

2: Pirl2:*

3: Pirl3:*

4: Pirl4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	24	100.0	548	1 I37577	islet cell antigen
2	24	100.0	922	1 S54342	protein-tyrosine-p
3	24	100.0	979	1 JC2349	protein-tyrosine-p
4	24	100.0	996	2 I48721	PTP 35 protein - m
5	9	37.5	248	2 T46903	hypothetical prote
6	9	37.5	1015	2 JC5062	phogrin precursor
7	9	37.5	1015	2 JC5263	transmembrane tyro
8	7	29.2	113	2 F90464	hypothetical prote
9	7	29.2	211	2 G90209	conserved hypothet
10	7	29.2	377	1 A48711	protein-tyrosine-p
11	7	29.2	397	2 T11777	phosphoglycerate t
12	7	29.2	397	2 A10806	phosphoglycerate t
13	7	29.2	398	2 I56540	protein-tyrosine-p
14	7	29.2	477	2 I38409	adenyl cyclase-a
15	7	29.2	519	2 T48498	hypothetical prote
16	7	29.2	582	2 A57068	protein-tyrosine-p
17	7	29.2	583	2 S17671	protein-tyrosine-p
18	7	29.2	668	2 T34317	protein-tyrosine-p
19	7	29.2	680	2 JC8052	protein tyrosine p
20	7	29.2	699	2 JC6132	protein-tyrosine-p
21	7	29.2	700	1 S2053	protein-tyrosine-p
22	7	29.2	711	1 S28391	protein-tyrosine-p
23	7	29.2	711	2 T45160	protein-tyrosine-p
24	7	29.2	756	2 S67433	hypothetical prote
25	7	29.2	796	1 JC1285	protein-tyrosine-p
26	7	29.2	802	1 A36065	protein-tyrosine-p
27	7	29.2	829	1 A47373	protein-tyrosine-p
28	7	29.2	839	2 C84697	hypothetical prote
29	7	29.2	1118	1 A49724	protein-tyrosine-p

ALIGNMENTS

RESULT 1

I37577

islet cell antigen 512 - human

N:Alternate names: islet cell autoantigen 3

C:Species: Homo sapiens (man)

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: I37577; S18121

R:Rabin, D.O.; Pleasick, S.M.; Shapiro, J.A.; Yoo-Warren, H.; Oles, J.; Hicks, J.M.; Gold

J. Immunol. 152, 3183-3188, 1994

A:Title: Islet cell antigen 512 is a diabetes-specific islet autoantigen related to prot

A:Reference number: I37577; MUID:94194080; PMID:8144912

A:Accession: I37577

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-548 <RES>

A:Cross-references: EMBL:X62899; NID:g32612; PIDN:CAA44688.1; PID:g32613

C:Genetics:

A:Gene: GDB:ICA3

A:Cross-references: GDB:385512

C:Superfamily: protein-tyrosine-phosphatase, receptor type N; protein-tyrosine-phosphata

C:Keywords: phosphoprotein; phosphoric monoester hydrolase; transmembrane protein; tyros

F:196-212/Domain: transmembrane #status predicted <TMN>

F:346-548/Domain: protein-tyrosine-phosphatase homology #status atypical <PTP>

F:521/Active site: Cys (phosphocysteine intermediate) #status predicted

Query Match 100.0%; Score 24; DB 1; Length 548;
Best Local Similarity 100.0%; Pred. NO. 6.3e-17; Mismatches 0; Indels 0; Gaps 0;
Matches 24; Conservative 0

Qy 1 KLKVSSPSRSDYINASPIIEHDP 24

Db 364 KLKVSSPSRSDYINASPIIEHDP 387

RESULT 2

S54342

protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type N precursor - rat

N:Alternate names: protein-tyrosine-phosphatase IA-2

C:Species: Rattus norvegicus (Norway rat)

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000

C:Accession: S54342

R:Kambayashi, Y.; Takahashi, K.; Bardhan, S.; Inagami, T.

Biochem. J. 306, 331-335, 1995

A:Title: Cloning and expression of protein tyrosine phosphatase-like protein derived fro

A:Reference number: S54342; MUID:95194305; PMID:7887886

A:Accession: S54342

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-922 <RAM>

A:Cross-references: GB:D38222; NID:gl504061; PIDN:BA07397.1; PID:g808890

C:Superfamily: protein-tyrosine-phosphatase, receptor type N; protein-tyrosine-phosphata

probable protein-t
protein-tyrosine-p
protein-tyrosine-p
leucocyte common a
protein-tyrosine-p
protein-tyrosine-p
protein-tyrosine-p
protein-tyrosine-p
protein-tyrosine-p
protein-tyrosine-p
leukocyte antigen-
leukocyte antigen-
protein-tyrosine-p
protein-tyrosine-p

30 7 29.2 1200 2 T43148
31 7 29.2 1231 2 S53089
32 7 29.2 1237 2 A54080
33 7 29.2 1290 2 A56493
34 7 29.2 1301 1 A41622
35 7 29.2 1496 1 A48758
36 7 29.2 1499 2 I50212
37 7 29.2 1501 2 I58148
38 7 29.2 1585 2 T19121
39 7 29.2 1691 1 D54689
40 7 29.2 1863 2 S46217
41 7 29.2 1894 2 C54689
42 7 29.2 1897 1 TDHULK
43 7 29.2 1898 2 S46216
44 7 29.2 1907 2 S50893
45 7 29.2 1912 2 A56178

C;Keywords: phosphoprotein; phosphoric monoester hydrolase; transmembrane protein; tyrosine phosphatase

F;525-541/Domain: transmembrane #status predicted <TM>

F;677-901/Domain: protein-tyrosine-phosphatase homology <PTP>

F;852/Active site: Cys (phosphocysteine intermediate) #status predicted

F;858/Binding site: substrate phosphate (Arg) #status predicted

Query Match 100.0%; Score 24; DB 1; Length 922;

Best Local Similarity 100.0%; Pred. No. 9.9e-17; Indels 0; Gaps 0;

Matches 24; Conservative 0; Mismatches 0;

QY 1 KLVKVESSPSRSDYINASPIIEHDP 24
 |||||
 Db 695 KLVKVESSPSRSDYINASPIIEHDP 718

RESULT 3

JC2349

protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type N precursor - mouse

N;Alternate names: protein-tyrosine-phosphatase IA-2

C;Species: Mus musculus (house mouse)

C;Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 09-Jul-2004

C;Accession: J02349; S40291; I48721

R;Ilu, J.; Notkins, A.L.; Lan, M.S.

Biochem. Biophys. Res. Commun. 204, 930-936, 1994

A;Title: Isolation, sequence and expression of a novel mouse brain cDNA, mIA-2, and its

A;Reference number: JC2349; MUID:95071416; PMID:7980563

A;Accession: JC2349

A;Molecule type: mRNA

A;Residues: 1-979 <LU>

A;Cross-references: UNIPROT:Q60673; EMBL:U11812; NID:G563737; PIDN:AAAS2102.1; PID:G563737

A;Experimental source: brain

R;Hendriks, W.; Brugman, C.; Zeeuwen, P.; Schepens, J.; Wieringa, B.

submitted to the EMBL Data Library, June 1993

A;Description: Assessment of the expression levels of murine protein-tyrosine phosphatases

A;Reference number: S40280

A;Accession: S40291

A;Molecule type: mRNA

A;Residues: 801-837, 839-852, 'T', 854-907 <HEN>

A;Cross-references: EMBL:Z23060; NID:G438157; PIDN:CAAR0595.1; PID:G438158

R;Magistrelli, G.; Covinin, N.; Mosca, M.; Lippoli, G.; Isacchi, A.

Biochem. Biophys. Res. Commun. 217, 154-161, 1995

A;Title: Expression of PTP35, the murine homologue of the PTP2 se-related sequences 1A-2

A;Reference number: I48721; MUID:96095652; PMID:8526904

A;Accession: I48721

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 'MPCARGRESSEPR', 1-165, 'GDGAGA', 170-362, 'L', 364-614, 'V', 616-674, 'T', 676-858

A;Cross-references: EMBL:X74438; NID:G1089901; PIDN:CAAS2453.1; PID:G1089902

C;Superfamily: protein-tyrosine-phosphatase, receptor type N; protein-tyrosine-phosphatase

C;Keywords: brain; glycoprotein; phosphoprotein; phosphoric monoester hydrolase; transmembrane

F;1-37/Domain: signal sequence #status predicted <SIG>

F;38-979/Product: protein-tyrosine-phosphatase, receptor type N #status predicted <NAT>

F;38-576/Domain: extracellular #status predicted <EXT>

F;577-598/Domain: transmembrane #status predicted <TM>

F;696-979/Domain: intracellular #status predicted <INT>

F;734-958/Domain: protein-tyrosine-phosphatase homology <PTP2>

F;506,524/Binding site: carbonylate (Asn) (covalent) #status predicted

F;909/Active site: Cys (phosphocysteine intermediate) #status predicted

F;915/Binding site: substrate phosphate (Arg) #status predicted

Query Match 100.0%; Score 24; DB 1; Length 979;

Best Local Similarity 100.0%; Pred. No. 1e-16;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVKVESSPSRSDYINASPIIEHDP 24
 |||||
 Db 752 KLVKVESSPSRSDYINASPIIEHDP 775

RESULT 4

I48721

PTP 35 protein - mouse

C;Species: Mus musculus (house mouse)

C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 23-Jul-1999

C;Accession: I48721

R;Magistrelli, G.; Covinin, N.; Mosca, M.; Lippoli, G.; Isacchi, A.

Biochem. Biophys. Res. Commun. 217, 154-161, 1995

A;Title: Expression of PTP35, the murine homologue of the PTP2 se-related sequences 1A-2

A;Reference number: I48721; MUID:96095652; PMID:8526904

A;Accession: I48721

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-996 <RES>

A;Cross-references: EMBL:X74438; NID:G1089901; PIDN:CAAS2453.1; PID:G1089902

C;Superfamily: protein-tyrosine-phosphatase, receptor type N; protein-tyrosine-phosphatase

C;Keywords: phosphoprotein

F;751-975/Domain: protein-tyrosine-phosphatase homology <PTP2>

F;926/Active site: Cys (phosphocysteine intermediate) #status predicted

F;932/Binding site: substrate phosphate (Arg) #status predicted

Query Match 100.0%; Score 24; DB 2; Length 996;

Best Local Similarity 100.0%; Pred. No. 1.1e-16;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVKVESSPSRSDYINASPIIEHDP 24
 |||||
 Db 769 KLVKVESSPSRSDYINASPIIEHDP 792

RESULT 5

T46903

hypothetical protein DKFZP761A0712.1 - human (fragment)

C;Species: Homo sapiens (man)

C;Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 09-Jul-2004

C;Accession: T46903

R;Ansorge, W.; Winkner, U.; Mewes, H.W.; Weil, B.; Wiemann, S.

submitted to the Protein Sequence Database, February 2000

A;Reference number: Z24134

A;Accession: T46903

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-248 <AAA>

A;Cross-references: UNIPROT:Q9NSR5; EMBL:AL157451

A;Experimental source: adult amygdala; clone DKFZP761A0712

C;Genetics:

A;Note: DKFZP761A0712.1

C;Superfamily: protein-tyrosine-phosphatase, receptor type N; protein-tyrosine-phosphatase

Query Match 37.5%; Score 9; DB 2; Length 248;

Best Local Similarity 100.0%; Pred. No. 0.057;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 SDYINASPI 19
 |||||
 Db 31 SDYINASPI 39

RESULT 6

JC5062

phogrin precursor - human

N;Contents: protein-tyrosine-phosphatase (EC 3.1.3.48) receptor type

C;Species: Homo sapiens (man)

C;Date: 31-Jan-1997 #sequence_revision 31-Jan-1997 #text_change 09-Jul-2004

C;Accession: JC5062; JC5263; T46903

R;Kawasaki, E.; Hutton, J.C.; Eisenbarth, G.S.

Biochem. Biophys. Res. Commun. 227, 440-447, 1996

A;Title: Molecular cloning and characterization of the human transmembrane protein tyrosine phosphatase

A;Reference number: JC5062; MUID:97032784; PMID:8878534

A;Contents: islet

A;Accession: JC5062

A;Molecule type: mRNA

A;Residues: 1-1015 <KAW>

A;Cross-references: UNIPROT:Q92932; GB:U66702; NID:G1620663; PIDN:AAAS0742.1; PID:G1620663

R;Smith, P.D.; Barker, K.T.; Wang, J.; Lu, Y.J.; Shipley, J.; Crompton, M.R.

Biochem. Biophys. Res. Commun. 229, 402-411, 1996

A;Title: ICARF, a novel member of a new family of transmembrane, tyrosine phosphatase-11

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 28, 2004, 19:58:09 ; Search time 2727.04 Seconds
(without alignments)
320.697 Million cell updates/sec

Title: US-10-783-095-12
Perfect score: 24
Sequence: 1 KLVESPSPSDYINASPPIEHPD 24

Scoring table:
OLIGO
Xgapop 60.0, Xgapext 60.0
Ygapop 60.0, Ygapext 60.0
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 32822875 seqs, 18219865908 residues

Word size: 1

Total number of hits satisfying chosen parameters: 65641363

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US10783095/runat_26102004_185454_25046/app_query.fasta_1.796
-DB=EST -QPMF=fastap -SUFFIX=oli.rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODBS=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10783095 @CGN_1_1_9525 @runat_26102004_185454_25046 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -Fgapop=6
-Fgapext=7 -Ygapop=60 -Ygapext=60 -DELOP=6 -DELEXT=7

Database : EST.*

1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_gss1.*
9: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	24	100.0	330	2	BB868512
2	24	100.0	357	2	AW670201
3	24	100.0	397	6	CB772195
4	24	100.0	404	6	CB811493
5	24	100.0	406	6	CB807662
6	24	100.0	426	6	CB895128
7	24	100.0	455	6	CB741139
8	24	100.0	477	6	CA541852
9	24	100.0	493	5	BQ559340

10	24	100.0	547	5	BQ831314
11	24	100.0	565	2	BF199250
12	24	100.0	583	4	BM507225
13	24	100.0	584	1	AI155087
14	24	100.0	600	4	BG799991
15	24	100.0	622	1	A260000
16	24	100.0	708	6	CB058651
17	24	100.0	711	7	CK471842
18	24	100.0	728	5	BQ745175
19	24	100.0	845	2	BF311980
20	24	100.0	846	2	BF304197
21	24	100.0	857	2	BF207292
22	24	100.0	912	5	BX433415
23	24	100.0	918	4	BG968308
24	24	100.0	926	2	BE799090
25	24	100.0	1071	4	BM805338
26	24	100.0	2900	9	AY39488
27	24	100.0	2940	9	AY39488
28	24	100.0	3515	3	AK076223
29	23	95.8	629	2	BF347251
30	22	91.7	401	6	CB698881
31	22	91.7	505	7	CK413735
32	20	83.3	917	6	CD517974
33	19	79.2	490	6	CB725836
34	19	79.2	705	9	EX981003
35	18	75.0	785	2	BE260974
36	17	70.8	673	2	BE786932
37	17	70.8	676	6	CF126653
38	15	62.5	829	4	BI199614
39	15	62.5	3556	3	AK041296
40	13	54.2	940	6	CD556898
41	12	50.0	422	6	CB763347
42	12	50.0	756	5	BM944030
43	11	45.8	297	7	CK690436
44	11	45.8	649	6	CD215646
45	10	41.7	661	1	AU168769

ALIGNMENTS

BB868512 330 bp mRNA linear EST 27-NOV-2001
BB868512 RIKEN full-length enriched, 16 days neonate male
diencephalon Mus musculus cDNA clone G63005N01 5', mRNA sequence.
BB868512
BB868512.1 GI:17114722
EST.*
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 330)

REFERENCE

AUTHORS

ORGANISM

KEYWORDS

ACCESSION

VERSION

LOCUS

DEFINITION

RESULT 1

BB868512

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (2001)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-resgsc.riken.jp, URL: http://genome.gsc.riken.jp/

RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al. 2001)

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* 10 (10), 1617-1630 (2000)

Wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)

Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

source

Location/Qualifiers

1..330
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="G630005N01"
/sex="male"

/tissue_type="diencephalon"
/dev_stages="16 days neonate"
/clone_lib="RIKEN full-length enriched, 16 days neonate male diencephalon"

ORIGIN

Alignment Scores:
Pred. No.: 2,898-14 Length: 330
Score: 24.00 Matches: 24
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 2 Gaps: 0

US-10-783-095-12 (1-24) x BB668512 (1-330)

Qy 1 LysLeuLysValGluSerProSerArgSerAspTyrIleAsnAlaSerProIlelle 20
|||
Db 222 AAGCTGAAGTGGAGAGCAGCCCTTCTCGAGTGATTACATCAGCCAGCCCATCATC 281
|||

Qy 21 GluHisAspPro 24
|||

Db 282 GAGCATGACCCCT 293
|||

RESULT 2

AW670201 357 bp mRNA linear EST 25-APR-2001
LOCUS 114090 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.

DEFINITION AW670201

ACCESSION AW670201

VERSION AW670201.1 GI:7526715

KEYWORDS EST.

SOURCE Bos taurus (cow)

ORGANISM Bos taurus

REFERENCE 1 (bases 1 to 357)

AUTHORS Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T., Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G., Perlea, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and Keele, J.W.

TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle

JOURNAL *Genome Res.* 11 (4), 626-630 (2001)

MEDLINE

PUBMED

COMMENT

21180013

11282978

Contact: Smith TPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smith@meat.mars.usda.gov

Single pass sequencing. Bases called and alt trimmed with phred

v0.980904.8. Vector identified by cross_match with the -minscore 18

and -minmatch 12 options.

PCR Primers

FORWARD: AGGAACACAGCTATGACCAT

BACKWARD: GTTTCCAGTCAGCAGC

Plate: 111 row: E column: 24

Seq primer: ATTAGGTGACACATATAG.

Location/Qualifiers

1..357

/organism="Bos taurus"

/mol_type="mRNA"

/db_xref="taxon:9913"

/tissue_type="pooled"

/lab_host="DH103"

/clone_lib="MARC 1BOV"

/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;

Library made from pooled tissue from lymph node, ovary,

fat, hypothalamus, and pituitary."

ORIGIN

Alignment Scores:
Pred. No.: 3,128-14 Length: 357
Score: 24.00 Matches: 24
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 2 Gaps: 0

US-10-783-095-12 (1-24) x AW670201 (1-357)

Qy 1 LysLeuLysValGluSerProSerArgSerAspTyrIleAsnAlaSerProIlelle 20
|||

Db 106 AAGCTGAAGTGGAGAGCAGCCCTTCTCGAGTGATTACATCAGCCAGCCCATATT 165
|||

Qy 21 GluHisAspPro 24
|||

Db 166 GAGCAGACCCCT 177
|||

RESULT 3

CB772195

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 397)

Angen EST Program.

Angen Rat EST Program

Unpublished (2003)

Contact: Dan Fitzpatrick

Angen, Inc

One Angen Center Drive, Thousand Oaks, CA 91320-1799, USA

Tel: 805 447-4881

Plate: 00011 row: d column: 3.

Location/Qualifiers

1..397

/organism="Rattus norvegicus"

/mol_type="mRNA"

CB772195 397 bp mRNA linear EST 16-MAY-2003
ANGNUC.TRPA3-00011-d3-A trpa3 (10298) Rattus norvegicus cDNA clone
trpa3-00011-d3 5', mRNA sequence.

CB772195

CB772195.1 GI:29860586

EST.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 397)

Angen EST Program.

Angen Rat EST Program

Unpublished (2003)

Contact: Dan Fitzpatrick

Angen, Inc

One Angen Center Drive, Thousand Oaks, CA 91320-1799, USA

Tel: 805 447-4881

Plate: 00011 row: d column: 3.

Location/Qualifiers

1..397

/organism="Rattus norvegicus"

/mol_type="mRNA"

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2004, 10:57:30 ; Search time 60.7229 Seconds
(without alignments)
227.410 Million cell updates/sec

Title: US-10-783-095-12
Perfect score: 24
Sequence: 1 KLVESPSPRSYINASPIIHDP 24

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0
Searched: 1825181 seqs, 575374646 residues
Word size : 0

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Uniprot 02:.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	24	100.0	358	2 Q6P3E6	Q6P3E6 mus musculus
2	24	100.0	358	2 AAH64020	AAH64020 mus musculus
3	24	100.0	591	2 Q7KZS4	Q7KZS4 homo sapien
4	24	100.0	811	2 Q961A0	Q961A0 homo sapien
5	24	100.0	811	2 AAH07713	AAH07713 homo sapi
6	24	100.0	873	2 Q925J6	Q925J6 mus musculus
7	24	100.0	876	2 Q925J7	Q925J7 mus musculus
8	24	100.0	950	2 Q6NSL1	Q6NSL1 homo sapien
9	24	100.0	979	1 PTPN_BOVIN	AAH70053
10	24	100.0	979	1 PTPN_HUMAN	AAH70053
11	24	100.0	979	1 PTPN_MOUSE	Q16849
12	24	100.0	983	1 PTPN_RAT	Q60673
13	24	100.0	983	1 PTPN_RAT	Q63259
14	24	100.0	985	2 Q6NVE2	Q6NVE2 mus musculus
15	24	100.0	985	2 AAH68165	AAH68165 mus musculus
16	9	37.5	248	2 Q9NSR5	Q9NSR5 mus musculus
17	9	37.5	986	2 Q9V4I9	Q9V4I9 homo sapien
18	9	37.5	986	2 Q9V4I6	Q9V4I6 homo sapien
19	9	37.5	1001	1 PTPX_MOUSE	P80560
20	9	37.5	1004	1 PTPX_RAT	Q63475
21	9	37.5	1013	1 PTPX_MACNE	Q02695
22	9	37.5	1015	1 PTPX_HUMAN	Q92932
23	9	37.5	1042	2 Q9V4F8	Q9V4F8 homo sapien
24	8	33.3	571	2 Q7S5A7	Q7S5A7 neurospora
25	8	33.3	571	2 CAP05891	CAI05891 neurospor
26	7	29.2	56	2 Q6UDH8	Q6UDH8 canis fami
27	7	29.2	62	2 Q90950	Q90950 gallus gall
28	7	29.2	63	2 Q90949	Q90949 gallus gall
29	7	29.2	63	2 Q90951	Q90951 gallus gall
30	7	29.2	113	2 Q97UX5	Q97UX5 sulfolobus
31	7	29.2	211	2 Q97ZQ9	Q97ZQ9 sulfolobus

32 7 29.2 229 2 Q845L1 Q845L1 bacillus me
33 7 29.2 240 2 Q9USB5 Q9USB5 schizosacch
34 7 29.2 297 2 Q63476 Q63476 rattus norv
35 7 29.2 352 2 Q8EY48 Q8EY48 debaryomyce
36 7 29.2 360 2 Q7SE42 Q7SE42 ashbya goss
37 7 29.2 360 2 AA50599 AA50599 ashbya go
38 7 29.2 374 2 Q8DFJ6 Q8DFJ6 dictyosteli
39 7 29.2 377 1 PTP2_DICDI PTP2_DICDI
40 7 29.2 380 1 CXA7_BRARE CXA7_BRARE
41 7 29.2 380 2 Q803S5 Q803S5 brachydanio
42 7 29.2 383 2 Q8MTN0 Q8MTN0 culicoides
43 7 29.2 385 2 Q7MM06 Q7MM06 vibrio vuln
44 7 29.2 397 1 PGTCSALTY P37591 salmonella
45 7 29.2 398 2 Q62604 Q62604 rattus norv

ALIGNMENTS

RESULT 1
Q6P3E6
ID Q6P3E6 PRELIMINARY; PRT; 358 AA.
AC Q6P3E6;
DT 05-JUL-2004 (Tremblrel. 27, Created)
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)
DE Ptpn protein (Fragment).
DE GN Name=Ptpn;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Colon;
RX MEDLINE=2238257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg S., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield V.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Colon;
RX Strausberg R.;
RL Submitted (DSC-2003) to the EMBL/GenBank/DBJ databases.
EMBL; BC064020; AAH64020.1; -;
DR InterPro; IPR003595; PTPC motif.
DR InterPro; IPR000387; TYR Phosphatase.
DR InterPro; IPR000242; Tyr.Pp.
DR Pfam; PF00102; Y_phosphatase; 1.
DR PRINTS; PR00700; PTRYPHPTASE.
DR SMART; SM00194; PTPC; 1.
DR SMART; SM00404; PTPC motif; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS00556; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS00555; TYR_PHOSPHATASE_PTP; 1.
KW Hydrolase.

FT NON TER 1 1
SQ SEQUENCE 358 AA; 40571 MW; 0152828A35992BF2 CRC64;
Query Match 100.0%; Score 24; DB 2; Length 358;
Best Local Similarity 100.0%; Pred. No. 2.8e-17; Indels 0; Gaps 0;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KLVKVESSPSRSDYINASPIIEHDP 24
Db 131 KLVKVESSPSRSDYINASPIIEHDP 154
RESULT 2
AAH64020 PRELIMINARY; PRT; 358 AA.
AC AAH64020;
DT 02-WAR-2004 (T-EMBLrel. 27, Created)
DT 02-WAR-2004 (T-EMBLrel. 27, Last sequence update)
DT 02-WAR-2004 (T-EMBLrel. 27, Last annotation update)
DE Ptpn protein (Fragment).
GN PTPRN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Colon;
RC MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Shenmen C.M., Schuler G.D.,
RA Klausner R.D., Collins P.S., Wagner L., Schaefer C.F., Bhat N.K.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
[2]
SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Colon;
RA Strausberg R.; (2003) to the EMBL/GenBank/DBJ databases.
DR ENBL; BC064020; AAH64020.1; -
FT NON TER 1
SQ SEQUENCE 358 AA; 40571 MW; 0152828A35992BF2 CRC64;
Query Match 100.0%; Score 24; DB 2; Length 358;
Best Local Similarity 100.0%; Pred. No. 2.8e-17; Indels 0; Gaps 0;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KLVKVESSPSRSDYINASPIIEHDP 24
Db 131 KLVKVESSPSRSDYINASPIIEHDP 154
RESULT 3
Q7KZS4 PRELIMINARY; PRT; 591 AA.
ID Q7KZS4;
AC Q7KZS4;
DT 05-JUL-2004 (T-EMBLrel. 27, Created)
DT 05-JUL-2004 (T-EMBLrel. 27, Last sequence update)

DT 05-JUL-2004 (T-EMBLrel. 27, Last annotation update)
DE Protein tyrosine phosphatase, receptor type, N.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RA Kalline N., Chen X., Rolfs A., Halleck A., Hines L., Eisenstein S.,
RA Koundinya M., Raphael J., Moreira D., Kelley T., Labaer J., Lin I.,
RA Phelan M., Farmer A.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR ENBL; BT006975; AAP35621.1; -
DR GO; GO:0004872; F_receptor_activity; IEA.
DR InterPro; IPR003595; PTPC_motif.
DR InterPro; IPR000387; TYR_phosphatase.
DR InterPro; IPR000242; Tyr_PP.
DR Pfam; PF00102; Y_phosphatase; 1.
DR PRINTS; PR00700; PRTYPHPTASE.
DR SMART; SM00194; PTPC; 1.
DR SMART; SM00404; PTPC_motif; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS00056; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS00055; TYR_PHOSPHATASE_PTP; 1.
KW Hydrolase; Receptor.
SQ SEQUENCE 591 AA; 64763 MW; 3AD00A90C03FF4F CRC64;
Query Match 100.0%; Score 24; DB 2; Length 591;
Best Local Similarity 100.0%; Pred. No. 4.5e-17; Indels 0; Gaps 0;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KLVKVESSPSRSDYINASPIIEHDP 24
Db 364 KLVKVESSPSRSDYINASPIIEHDP 387
RESULT 4
Q96IAO PRELIMINARY; PRT; 811 AA.
ID Q96IAO;
AC Q96IAO;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE PTPRN protein (fragment).
GN Name=PTPRN;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=Brain;
RC MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2004, 06:03:19 ; Search time 84.7229 Seconds
(without alignments)
101.620 Million cell updates/sec

Title: US-10-783-095-12

Perfect score: 123
Sequence: 1 KLKVSSPSRSDYINASPIIEHDP 24

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_23Sep04:*

- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001s:*
- 5: Geneseq2002s:*
- 6: Geneseq2003as:*
- 7: Geneseq2003bs:*
- 8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	123	100.0	294	2	AAY33360 Human isl
2	123	100.0	294	2	AAY49854 Human pan
3	123	100.0	294	2	AAY80486 Islet cel
4	123	100.0	294	3	AAY49329 Pancreat
5	123	100.0	379	3	AAY59351 Tyrosine
6	123	100.0	548	2	AAY33361 Human isl
7	123	100.0	548	2	AAY49855 Human pan
8	123	100.0	548	2	AAY80487 Islet cel
9	123	100.0	548	3	AAY49330 Pancreat
10	123	100.0	548	6	AAY49330 Human exp
11	123	100.0	548	6	ABU04788 Human exp
12	123	100.0	548	6	ABU04790 Human exp
13	123	100.0	548	6	ABU04792 Human exp
14	123	100.0	548	6	ABU04791 Human exp
15	123	100.0	591	6	ABU04783 Human exp
16	123	100.0	591	6	ABU04785 Human exp
17	123	100.0	916	6	ABU04795 Human exp
18	123	100.0	916	2	AAR45779 PTPase PT
19	123	100.0	916	6	ABU04786 Human exp
20	123	100.0	923	7	AD56165 Rat Prote
21	123	100.0	923	7	AD445276 Rat Prote
22	123	100.0	934	5	ABU06112 Human NS
23	123	100.0	961	2	AAR49039 PTPase PT
24	123	100.0	961	6	ABU04787 Human exp
25	123	100.0	978	8	AD060054 CRH signa
26	123	100.0	979	2	AAY06606 Human isl

26	123	100.0	979	4	AAG80205 Human aut
27	123	100.0	979	6	ABU04793 Human exp
28	123	100.0	979	6	ABU04784 Human exp
29	123	100.0	979	6	ABU04789 Human exp
30	123	100.0	979	6	ABU04794 Human exp
31	123	100.0	979	7	AD56167 Human Pro
32	123	100.0	979	7	AD56167 Human Pro
33	123	100.0	979	7	AD56167 Human Pro
34	123	100.0	979	7	AD56167 Human Pro
35	123	100.0	979	7	AD56167 Human Pro
36	123	100.0	979	7	AD56167 Human Pro
37	123	100.0	979	7	AD56167 Human Pro
38	114	92.7	22	3	ABB57041 Mouse isl
39	92	74.8	19	3	ABB57041 Mouse isl
40	91	74.0	298	2	ABB57041 Mouse isl
41	91	74.0	370	2	ABB57041 Mouse isl
42	91	74.0	376	2	ABB57041 Mouse isl
43	91	74.0	641	2	ABB57041 Mouse isl
44	91	74.0	818	2	ABB57041 Mouse isl
45	91	74.0	876	2	ABB57041 Mouse isl

ALIGNMENTS

RESULT 1
AAY33360
ID AAY33360 standard; protein; 294 AA.

XX AC AAY33360;
XX AC
XX 29-NOV-1999 (first entry)
XX DE Human islet cell antigen clone ICA-512 protein fragment.
XX KW Islet cell antigen; ICA; human; pancreas; antibody; IDDM; immunoassay;
XX KW insulin-dependent type I diabetes mellitus; diagnosis; immunoglobulin;
XX KW T-cell; B-cell; islet cell destruction; immunoreactivity; disease;
XX KW immune therapy; stimulating antigen; screening.
XX OS Homo sapiens.
XX PN US955345-A.
XX PD 21-SEP-1999.
XX PF 06-JUN-1995; 95US-00468576.
XX PR 17-FEB-1989; 89US-00312543.
XX PR 04-DEC-1989; 89US-00441703.
XX PR 14-JUN-1991; 91US-00715181.
XX PR 08-JUN-1992; 92US-00872646.
XX PR 05-MAY-1994; 94US-00239276.
XX PA (FARB) BAYER CORP FORMERLY MOLECULAR DIAGNOSTIC.
XX Rabin DU;
XX WPI; 1999-560494/47.
XX N-PSDB; AAZ09922.
XX Isolated DNA encoding antigens reactive with serum from diabetics useful
XX PT to develop products for the diagnosis and monitoring of insulin-dependent
XX PT diabetes mellitus.
XX Example; Col 85-87; 63pp; English.
XX CC This invention describes novel isolated and purified DNAs encoding
XX CC pancreatic islet cell antigens reactive with islet cell antibodies found
XX CC in sera of patients with insulin-dependent type I diabetes mellitus
XX CC (IDDM). The polypeptides of the invention are useful as immunoassay
XX CC reagents in the presymptomatic diagnosis of insulin-dependent type I
XX CC diabetes mellitus (IDDM). The ability of the polypeptides to bind the

CC antibody binding site on islet cell antibodies (ICAs) also confers
 CC utility in the binding or blocking of human immunoglobulin, T-cells, or B
 CC -cells involved in IDDM. The products can also be used to study the
 CC biological mechanisms involved in islet cell destruction and the
 CC appearance of ICA. The immunoreactivity profile with different antigens
 CC can provide diagnostically significant information concerning the nature
 CC of the disease, e.g. subtypes, the state of the disease, the proximity to
 CC onset of the disease, and the efficacy of therapy, e.g. immune therapy.
 CC The antigens can serve as stimulating antigens for T-cell culture.
 CC permitting significantly improved T-cell cloning, identification, and
 CC growth. The availability of large quantities of pure antigen enables the
 CC development of highly sensitive and specific immunoassays which can be
 CC used to screen the general population of presymptomatic IDDM or a
 CC predisposition to develop IDDM. This sequence represents the human islet
 CC cell antigen from clone ICA-512
 XX
 SQ Sequence 294 AA;

Query Match 100.0%; Score 123; DB 2; Length 294;
 Best Local Similarity 100.0%; Pred. No. 2.1e-11;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLVESPPSRSDYINASPIIEHDP 24
 |||||
 Db 110 KLVESPPSRSDYINASPIIEHDP 133

RESULT 2
 AAY49854
 ID AAY49854 standard; protein; 294 AA.
 XX
 AC AAY49854;
 XX
 DT 21-JAN-2000 (first entry)
 DE Human pancreatic islet cell antigen ICA-512 ATCC 40706 protein.
 XX
 KW Human; pancreatic islet cell antigen; ICA clone; IDDM; diabetes;
 KW insulin dependent type 1 diabetes mellitus; diagnosis; detection;
 KW immunoglobulin; T-cell; B-cell; antibody binding.
 XX
 OS Homo sapiens.
 XX
 PN US5981700-A.
 XX
 PD 09-NOV-1999.
 XX
 PF 06-JUN-1995; 95US-00468579.
 XX
 PR 17-FEB-1989; 89US-00312543.
 PR 04-DEC-1989; 89US-00441703.
 PR 14-JUN-1991; 91US-00715181.
 PR 08-JUN-1992; 92US-00872646.
 PR 05-MAY-1994; 94US-00239276.
 XX
 PA (FARB) BAYER CORP FORMERLY MOLECULAR DIAGNOSTIC.
 XX
 PI Rabin DU;
 XX
 XX WPI; 1999-633370/54.
 DR N-PSDB; AAZ32341.
 DR
 XX New pancreatic islet cell antigens, useful in the diagnosis of insulin-
 PT dependent (Type 1) diabetes mellitus.
 PT
 PS Claim 1; Col 83-86; 61pp; English.
 XX
 CC The present sequence represents a human pancreatic islet cell antigen
 CC (ICA) that binds with antibodies found in the sera of patients afflicted
 CC with insulin-dependent (Type 1) diabetes mellitus (IDDM). ICAs from the
 CC present invention are encoded by the DNA insert of a recombinant cloning
 CC vehicle selected from ATCC 40550 (AAZ32333, encoding AAY49847), ATCC
 CC 40553 (AAZ32334, encoding AAY49848), ATCC 40554 (AAZ32335, encoding

CC AAY49849), ATCC 40551 (AAZ32336, encoding AAY49850), ATCC 40552
 CC (AAZ32337, encoding AAY49851), ATCC 40703 (AAZ32338, encoding AAY49852),
 CC ATCC 40704 (AAZ32339, encoding AAY49853), ATCC 40705 (AAZ32340), ATCC
 CC 40706 (AAZ32341, encoding AAY49854) and ATCC 75030 (AAZ32342, encoding
 CC AAY49855). ICA proteins and their peptide fragments can be used in the
 CC diagnosis of IDDM and in detecting or blocking human immunoglobulin, T-
 CC cells or B-cells involved in IDDM
 XX
 SQ Sequence 294 AA;

Query Match 100.0%; Score 123; DB 2; Length 294;
 Best Local Similarity 100.0%; Pred. No. 2.1e-11;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLVESPPSRSDYINASPIIEHDP 24
 |||||
 Db 110 KLVESPPSRSDYINASPIIEHDP 133

RESULT 3
 AAW80486
 ID AAW80486 standard; protein; 294 AA.
 XX
 AC AAW80486;
 XX
 DT 01-FEB-1999 (first entry)
 DE Islet cell antibody antigen encoded by clone ICA-512.
 XX
 KW Pancreatic islet cell antibody antigen; ICA antigen; diagnosis;
 KW treatment; insulin dependent diabetes mellitus; IDDM; autoantibody;
 KW autoimmune disease; human.
 XX
 OS Homo sapiens.
 XX
 PN US5840836-A.
 XX
 PD 24-NOV-1998.
 XX
 PF 05-MAY-1994; 94US-00239276.
 XX
 PR 17-FEB-1989; 89US-00312543.
 PR 04-DEC-1989; 89US-00441703.
 PR 14-JUN-1991; 91US-00715181.
 PR 08-JUN-1992; 92US-00872646.
 XX
 PA (FARB) BAYER CORP.
 XX
 PI Rabin DU;
 XX
 XX WPI; 1999-034118/03.
 DR N-PSDB; AAV63560.
 DR
 XX Polypeptide antigens from pancreatic islet cells - useful for the
 PT detection and treatment of insulin dependent diabetes mellitus.
 PT
 PS Claim 1; Col 51-54; 47pp; English.
 XX
 CC The present sequence represents a pancreatic islet cell antibody (ICA)
 CC antigen. The proteins are used for the diagnosis and treatment of insulin
 CC dependent (Type 1) diabetes mellitus (IDDM). They are recognised by
 CC autoantibodies of the disease as IDDM is now thought to be an autoimmune
 CC disease. They were identified from probing of IDDM serum with human cDNA
 CC libraries expressing protein fragments
 XX
 SQ Sequence 294 AA;

Query Match 100.0%; Score 123; DB 2; Length 294;
 Best Local Similarity 100.0%; Pred. No. 2.1e-11;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLVESPPSRSDYINASPIIEHDP 24
 |||||
 Db 110 KLVESPPSRSDYINASPIIEHDP 133

RESULT 3
 AAW80486
 ID AAW80486 standard; protein; 294 AA.
 XX
 AC AAW80486;
 XX
 DT 01-FEB-1999 (first entry)
 DE Islet cell antibody antigen encoded by clone ICA-512.
 XX
 KW Pancreatic islet cell antibody antigen; ICA antigen; diagnosis;
 KW treatment; insulin dependent diabetes mellitus; IDDM; autoantibody;
 KW autoimmune disease; human.
 XX
 OS Homo sapiens.
 XX
 PN US5840836-A.
 XX
 PD 24-NOV-1998.
 XX
 PF 05-MAY-1994; 94US-00239276.
 XX
 PR 17-FEB-1989; 89US-00312543.
 PR 04-DEC-1989; 89US-00441703.
 PR 14-JUN-1991; 91US-00715181.
 PR 08-JUN-1992; 92US-00872646.
 XX
 PA (FARB) BAYER CORP.
 XX
 PI Rabin DU;
 XX
 XX WPI; 1999-034118/03.
 DR N-PSDB; AAV63560.
 DR
 XX Polypeptide antigens from pancreatic islet cells - useful for the
 PT detection and treatment of insulin dependent diabetes mellitus.
 PT
 PS Claim 1; Col 51-54; 47pp; English.
 XX
 CC The present sequence represents a pancreatic islet cell antibody (ICA)
 CC antigen. The proteins are used for the diagnosis and treatment of insulin
 CC dependent (Type 1) diabetes mellitus (IDDM). They are recognised by
 CC autoantibodies of the disease as IDDM is now thought to be an autoimmune
 CC disease. They were identified from probing of IDDM serum with human cDNA
 CC libraries expressing protein fragments
 XX
 SQ Sequence 294 AA;

Query Match 100.0%; Score 123; DB 2; Length 294;
 Best Local Similarity 100.0%; Pred. No. 2.1e-11;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLVESPPSRSDYINASPIIEHDP 24
 |||||
 Db 110 KLVESPPSRSDYINASPIIEHDP 133

RESULT 3
 AAW80486
 ID AAW80486 standard; protein; 294 AA.
 XX
 AC AAW80486;
 XX
 DT 01-FEB-1999 (first entry)
 DE Islet cell antibody antigen encoded by clone ICA-512.
 XX
 KW Pancreatic islet cell antibody antigen; ICA antigen; diagnosis;
 KW treatment; insulin dependent diabetes mellitus; IDDM; autoantibody;
 KW autoimmune disease; human.
 XX
 OS Homo sapiens.
 XX
 PN US5840836-A.
 XX
 PD 24-NOV-1998.
 XX
 PF 05-MAY-1994; 94US-00239276.
 XX
 PR 17-FEB-1989; 89US-00312543.
 PR 04-DEC-1989; 89US-00441703.
 PR 14-JUN-1991; 91US-00715181.
 PR 08-JUN-1992; 92US-00872646.
 XX
 PA (FARB) BAYER CORP.
 XX
 PI Rabin DU;
 XX
 XX WPI; 1999-034118/03.
 DR N-PSDB; AAV63560.
 DR
 XX Polypeptide antigens from pancreatic islet cells - useful for the
 PT detection and treatment of insulin dependent diabetes mellitus.
 PT
 PS Claim 1; Col 51-54; 47pp; English.
 XX
 CC The present sequence represents a pancreatic islet cell antibody (ICA)
 CC antigen. The proteins are used for the diagnosis and treatment of insulin
 CC dependent (Type 1) diabetes mellitus (IDDM). They are recognised by
 CC autoantibodies of the disease as IDDM is now thought to be an autoimmune
 CC disease. They were identified from probing of IDDM serum with human cDNA
 CC libraries expressing protein fragments
 XX
 SQ Sequence 294 AA;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2004, 10:22:24 : Search time 25.4458 Seconds
(without alignments)
62.550 Million cell updates/sec

Title: US-10-783-095-12
Perfect score: 123
Sequence: 1 KLVESPSRSDYINASPIIEHDP 24

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/prodata/1/aa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/aa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/aa/6A_COMB.pep:*
4: /cgn2_6/prodata/1/aa/6B_COMB.pep:*
5: /cgn2_6/prodata/1/aa/PCITUS_COMB.pep:*
6: /cgn2_6/prodata/1/aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	123	100.0	246	3 US-08-884-569A-4	Sequence 4, Appli
2	123	100.0	294	2 US-08-468-576B-18	Sequence 18, Appl
3	123	100.0	294	2 US-08-468-579B-18	Sequence 18, Appl
4	123	100.0	294	2 US-08-468-577B-18	Sequence 18, Appl
5	123	100.0	548	2 US-08-468-576B-19	Sequence 19, Appl
6	123	100.0	548	2 US-08-468-579B-19	Sequence 19, Appl
7	123	100.0	548	3 US-08-468-577B-19	Sequence 19, Appl
8	123	100.0	979	3 US-08-514-213A-2	Sequence 2, Appli
9	123	100.0	979	4 US-09-015-399-5	Sequence 5, Appli
10	114	92.7	22	4 US-09-552-802B-41	Sequence 41, Appl
11	92	74.8	19	4 US-09-552-802B-42	Sequence 42, Appl
12	91	74.0	298	3 US-08-811-481-4	Sequence 4, Appli
13	91	74.0	298	4 US-09-876-527-4	Sequence 4, Appli
14	91	74.0	376	3 US-08-811-481-7	Sequence 7, Appli
15	91	74.0	376	4 US-09-876-527-7	Sequence 7, Appli
16	91	74.0	641	3 US-08-811-481-2	Sequence 2, Appli
17	91	74.0	641	4 US-09-876-527-2	Sequence 2, Appli
18	91	74.0	818	3 US-08-811-481-22	Sequence 22, Appl
19	91	74.0	818	4 US-09-876-527-22	Sequence 22, Appl
20	91	74.0	969	2 US-08-548-159-1	Sequence 1, Appli
21	91	74.0	986	2 US-08-548-159-3	Sequence 3, Appli
22	91	74.0	1012	3 US-08-811-481-16	Sequence 16, Appl
23	91	74.0	1012	4 US-09-876-527-16	Sequence 16, Appl
24	82	65.7	17	4 US-09-552-802B-40	Sequence 40, Appl
25	81	65.9	246	3 US-08-884-569A-3	Sequence 3, Appli
26	81	65.9	723	2 US-08-548-159-5	Sequence 5, Appli
27	81	65.9	1001	3 US-08-884-569A-2	Sequence 2, Appli

28 59 48.0 12 4 US-09-552-802B-38 Sequence 38, Appli
29 59 48.0 12 4 US-09-552-802B-39 Sequence 39, Appli
30 55 44.7 304 4 US-09-552-802B-52 Sequence 52, Appli
31 55 44.7 304 4 US-09-270-767-46239 Sequence 46239, A
32 49 39.8 71 4 US-09-513-999C-5529 Sequence 5529, A
33 49 39.8 449 2 US-09-355-214-3 Sequence 3, Appli
34 49 39.8 456 2 US-08-819-013-1 Sequence 1, Appli
35 49 39.8 456 4 US-09-355-214-1 Sequence 1, Appli
36 49 39.8 861 4 US-09-538-092-809 Sequence 809, Appl
37 48 39.0 325 4 US-09-248-796A-14989 Sequence 14989, A
38 46 37.4 455 4 US-09-793-998-8 Sequence 8, Appli
39 46 37.4 587 3 US-08-899-578-2 Sequence 2, Appli
40 46 37.4 1267 4 US-09-543-681A-6130 Sequence 6130, A
41 45 36.6 134 4 US-09-252-991A-25959 Sequence 25959, A
42 45 36.6 254 2 US-08-685-992-23 Sequence 23, Appl
43 45 36.6 254 2 US-09-144-325-23 Sequence 23, Appl
44 44.5 36.2 1049 4 US-09-252-991A-17298 Sequence 17298, A
45 44 35.8 128 4 US-09-134-000C-3995 Sequence 3995, A

ALIGNMENTS

RESULT 1
US-08-884-569A-4
; Sequence 4, Application US/0884569A
; Patent No. 6399326
; GENERAL INFORMATION:
; APPLICANT: CHIANG, MING-KO
; APPLICANT: FLANAGAN, JOHN G.
; TITLE OF INVENTION: RECEPTOR TYROSINE PHOSPHATASE, AND USES RELATED THERETO
; FILE REFERENCE: HMV-020.01
; CURRENT APPLICATION NUMBER: US/08/884,569A
; PRIOR FILING DATE: 1997-06-27
; PRIOR APPLICATION NUMBER: 60/021,040
; PRIOR FILING DATE: 1996-07-02
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 4
; LENGTH: 246
; TYPE: PRT
; ORGANISM: Mus sp.
US-08-884-569A-4

Query Match 100.0%; Score 123; DB 3; Length 246;
Best Local Similarity 100.0%; Pred. No. 1.1e-12;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLVESPSRSDYINASPIIEHDP 24
Db 19 KLVESPSRSDYINASPIIEHDP 42

RESULT 2
US-08-468-576B-18
; Sequence 18, Application US/08468576B
; Patent No. 5955345
; GENERAL INFORMATION:
; APPLICANT: Rabin, Daniel
; TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
; TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Sprung Kramer Schaefer & Briscoe
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: System 7.5

```

; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,576B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/239,276
; FILING DATE: 05-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,646
; FILING DATE: 08-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/715,181
; FILING DATE: 14-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/441,703
; FILING DATE: 04-DEC-1989
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 294 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-468-576B-18

Query Match 100.0%; Score 123; DB 2; Length 294;
Best Local Similarity 100.0%; Pred. No. 1.4e-12;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVSSPSRSDYINASPIIHDHP 24
Db 110 KLVSSPSRSDYINASPIIHDHP 133

RESULT 3
US-08-468-579B-18
; Sequence 18, Application US/08468579B
; Patent No. 5981700
; GENERAL INFORMATION:
; APPLICANT: Rabin, Daniel
; TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sprung Kramer Schaefer & Briscoe
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: System 7.5
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,579B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/239,276
; FILING DATE: 05-MAY-1994
; APPLICATION NUMBER: US 07/872,646

```

```

; FILING DATE: 08-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/715,181
; FILING DATE: 14-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/441,703
; FILING DATE: 04-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/312,543
; FILING DATE: 17-FEB-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Kurt G. Briscoe
; REGISTRATION NUMBER: 33,141
; REFERENCE/DOCKET NUMBER: MDI 251.5-KGB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914) 332-1700
; TELEFAX: (914) 332-1844
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 294 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-468-579B-18

Query Match 100.0%; Score 123; DB 2; Length 294;
Best Local Similarity 100.0%; Pred. No. 1.4e-12;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVSSPSRSDYINASPIIHDHP 24
Db 110 KLVSSPSRSDYINASPIIHDHP 133

RESULT 4
US-08-468-577B-18
; Sequence 18, Application US/08468577B
; Patent No. 6001804
; GENERAL INFORMATION:
; APPLICANT: Rabin, Daniel
; TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sprung Kramer Schaefer & Briscoe
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: System 7.5
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,577B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/239,276
; FILING DATE: 05-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,646
; FILING DATE: 08-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/715,181
; FILING DATE: 14-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/441,703
; FILING DATE: 04-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/312,543
; FILING DATE: 17-FEB-1989

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2004, 10:54:39 ; Search time 313.735 Seconds
(without alignments)
24.802 Million cell updates/sec

Title: US-10-783-095-12

Perfect score: 123
Sequence: 1 KLVESPSRSDYINASPIIEHDP 24

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1370721 seqs, 324215800 residues

Total number of hits satisfying chosen parameters: 1370721

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	123	100.0	979	14 US-10-038-686-3	Sequence 3, Appli
2	114	92.7	22	15 US-10-378-479-41	Sequence 41, Appl
3	92	74.8	19	15 US-10-378-479-42	Sequence 42, Appl
4	91	74.0	298	9 US-09-876-527-4	Sequence 4, Appli
5	91	74.0	298	14 US-10-124-089-4	Sequence 4, Appli
6	91	74.0	376	9 US-09-876-527-7	Sequence 7, Appli
7	91	74.0	376	14 US-10-124-089-7	Sequence 7, Appli
8	91	74.0	641	9 US-09-876-527-2	Sequence 2, Appli
9	91	74.0	641	14 US-10-124-089-2	Sequence 2, Appli
10	91	74.0	818	9 US-09-876-527-22	Sequence 22, Appl
11	91	74.0	818	14 US-10-124-089-22	Sequence 22, Appl
12	91	74.0	1012	9 US-09-876-527-16	Sequence 16, Appl
13	91	74.0	1012	14 US-10-124-089-16	Sequence 16, Appl

14	82	66.7	17	15	US-10-378-479-40	Sequence 40, Appl
15	59	48.0	12	15	US-10-378-479-38	Sequence 38, Appl
16	59	48.0	12	15	US-10-378-479-39	Sequence 39, Appl
17	55	44.7	11	15	US-10-378-479-52	Sequence 52, Appl
18	49	39.8	213	16	US-10-437-963-115384	Sequence 115384,
19	48	39.0	66	15	US-10-424-599-189618	Sequence 189618,
20	48	39.0	535	14	US-10-032-585-7255	Sequence 7255, Ap
21	48	39.0	609	14	US-10-177-478-9	Sequence 9, Appli
22	46	37.4	112	14	US-10-104-047-2511	Sequence 2511, Ap
23	46	37.4	432	16	US-10-437-963-196840	Sequence 196840,
24	46	37.4	455	9	US-09-793-998-8	Sequence 8, Appli
25	46	37.4	455	15	US-10-429-893-8	Sequence 8, Appli
26	46	37.4	579	14	US-10-369-493-6573	Sequence 6573, Ap
27	46	37.4	587	16	US-10-687-732-25	Sequence 25, Appl
28	45.5	37.0	70	15	US-10-424-599-211903	Sequence 211903,
29	45.5	37.0	390	16	US-10-437-963-131675	Sequence 131675,
30	45	36.6	66	16	US-10-437-963-111307	Sequence 111307,
31	45	36.6	205	16	US-10-437-963-188792	Sequence 188792,
32	45	36.6	312	9	US-09-788-626-24	Sequence 24, Appl
33	45	36.6	312	14	US-10-369-493-15261	Sequence 15261, A
34	45	36.6	312	15	US-10-389-566-2435	Sequence 2435, Ap
35	45	36.6	361	15	US-10-425-114-64408	Sequence 64408, A
36	45	36.6	370	15	US-10-425-114-62602	Sequence 62602, A
37	45	36.6	455	15	US-10-425-114-69763	Sequence 69763, A
38	45	36.6	738	15	US-10-282-122A-77480	Sequence 77480, A
39	45	36.6	1333	15	US-10-621-758A-12	Sequence 12, Appl
40	45	36.6	1333	16	US-10-663-208A-12	Sequence 12, Appl
41	45	36.6	1333	16	US-10-646-301A-12	Sequence 12, Appl
42	45	36.6	1333	16	US-10-736-769-12	Sequence 12, Appl
43	45	36.6	3647	16	US-10-437-963-119793	Sequence 119793,
44	44	35.8	95	9	US-09-925-300-1824	Sequence 1824, Ap
45	44	35.8	201	16	US-10-767-701-32651	Sequence 32651, A

ALIGNMENTS

RESULT 1
US-10-038-686-3 ; Sequence 3, Application US/10038686
; Publication No. US20030045467A1
; GENERAL INFORMATION:
; APPLICANT: Orban, Tihamer
; TITLE OF INVENTION: AUTOANTIGEN VACCINE
; FILE REFERENCE: 10276-067001
; CURRENT APPLICATION NUMBER: US/10/038,686
; PRIOR FILING DATE: 2002-05-31
; PRIOR APPLICATION NUMBER: 60/260,068
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 979
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-038-686-3

Query Match 100.0%; Score 123; DB 14; Length 979;
Best Local Similarity 100.0%; Pred. No. 6.6e-10; Indels 0; Gaps 0;
Matches 24; Conservative 0; Mismatches 0

Qy 1 KLVESPSRSDYINASPIIEHDP 24
Db 752 KLVESPSRSDYINASPIIEHDP 775
RESULT 2
US-10-378-479-41
; Sequence 41, Application US/10378479
; Publication No. US20040006202A1
; GENERAL INFORMATION:
; APPLICANT: Peakman, Mark
; APPLICANT: Chiccz, Roman M.

```

; TITLE OF INVENTION: PEPTIDE EPITOPES RECOGNIZED BY DISEASE PROMOTING
; TITLE OF INVENTION: CD4+ T LYMPHOCYTES
; FILE REFERENCE: 08191-009002
; CURRENT APPLICATION NUMBER: US/10/378,479
; CURRENT FILING DATE: 2003-03-03
; PRIOR APPLICATION NUMBER: US/09/552,802B
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 09/295,868
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: US 60/130,355
; PRIOR FILING DATE: 1999-04-21
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 4.0.
; SEQ ID NO 41
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-378-479-41

```

```

Query Match          92.7%; Score 114; DB 15; Length 22;
Best Local Similarity 100.0%; Pred. No. 2.1e-10;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      3 KVESPPSRSDYINASPIIHDP 24
      |||||
Db      1 KVESPPSRSDYINASPIIHDP 22

```

```

RESULT 3
US-10-378-479-42
; Sequence 42, Application US/10378479
; Publication No. US20040006202A1
; GENERAL INFORMATION:
; APPLICANT: Peakman, Mark
; APPLICANT: Chiciz, Roman M.
; TITLE OF INVENTION: PEPTIDE EPITOPES RECOGNIZED BY DISEASE PROMOTING
; TITLE OF INVENTION: CD4+ T LYMPHOCYTES
; FILE REFERENCE: 08191-009002
; CURRENT APPLICATION NUMBER: US/10/378,479
; CURRENT FILING DATE: 2003-03-03
; PRIOR APPLICATION NUMBER: US/09/552,802B
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 09/295,868
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: US 60/130,355
; PRIOR FILING DATE: 1999-04-21
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 42
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-378-479-42

```

```

Query Match          74.8%; Score 92; DB 15; Length 19;
Best Local Similarity 100.0%; Pred. No. 4.1e-07;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      2 KVESPPSRSDYINASPII 20
      |||||
Db      1 KVESPPSRSDYINASPII 19

```

```

RESULT 4
US-09-876-527-4
; Sequence 4, Application US/09876527
; Patent No. US20020102616A1
; GENERAL INFORMATION:
; APPLICANT: Kindsvogel, Wayne
; Jelinek, Laura J.
; Sheppard, Paul O.
; Hagopian, William A.
; LaGasse, James M.

```

```

; TITLE OF INVENTION: ISLET CELL ANTIGEN 1851
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/876,527
; FILING DATE: 07-Jun-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/811,481
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Lingenfelter, Susan
; REGISTRATION NUMBER: P-41,156
; REFERENCE/DOCKET NUMBER: 95-36
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6675
; TELEFAX: 206-442-6678
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 298 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-876-527-4

```

```

Query Match          74.0%; Score 91; DB 9; Length 298;
Best Local Similarity 73.9%; Pred. No. 1.4e-05;
Matches 17; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

```

```

QY      2 LKVESPPSRSDYINASPIIHDP 24
      |||||
Db      150 LKAEENSHSHDYINASPIINDHP 172

```

```

RESULT 5
US-10-124-089-4
; Sequence 4, Application US/10124089
; Publication No. US20030166067A1
; GENERAL INFORMATION:
; APPLICANT: Kindsvogel, Wayne
; Jelinek, Laura J.
; Sheppard, Paul O.
; Hagopian, William A.
; LaGasse, James M.

```

```

; TITLE OF INVENTION: ISLET CELL ANTIGEN 1851
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0

```

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2004, 10:55:55 ; Search time 77.5904 Seconds
(without alignments)
129.454 Million cell updates/sec

Title: US-10-783-095-11

Perfect score: 28
Sequence: 1 LAKEWALCAYQAEPNTCAQEGNIK 28

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 2002273 seqs, 358729299 residues

Word size: 0

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: A_Geneseq_23Sep04:*

- 1: Genesecp1980s:*
- 2: Genesecp1990s:*
- 3: Genesecp2000s:*
- 4: Genesecp2001s:*
- 5: Genesecp2002s:*
- 6: Genesecp2003as:*
- 7: Genesecp2003bs:*
- 8: Genesecp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	28	100.0	294	2	AAY33360 Human isl
2	28	100.0	294	2	AAY49854 Human pan
3	28	100.0	294	2	AAY80486 Islet cel
4	28	100.0	294	3	AAY49329 Pancreat
5	28	100.0	279	3	AAY59351 Tyrosine
6	28	100.0	548	2	AAY33361 Human isl
7	28	100.0	548	2	AAY49855 Human pan
8	28	100.0	548	2	AAY80487 Islet cel
9	28	100.0	548	3	AAY49330 Pancreat
10	28	100.0	548	6	ABU04788 Human exp
11	28	100.0	548	6	ABU04790 Human exp
12	28	100.0	548	6	ABU04792 Human exp
13	28	100.0	548	6	ABU04791 Human exp
14	28	100.0	591	6	ABU04783 Human exp
15	28	100.0	591	6	ABU04785 Human exp
16	28	100.0	591	6	ABU04795 Human exp
17	28	100.0	923	7	AD56165 Rat Prote
18	28	100.0	923	7	AD45276 Rat Prote
19	28	100.0	934	5	AB506112 Human NS
20	28	100.0	979	4	AAY05606 Human isl
21	28	100.0	979	4	AAG50205 Human aut
22	28	100.0	979	6	ABU04793 Human exp
23	28	100.0	979	6	ABU04784 Human exp
24	28	100.0	979	6	ABU04789 Human exp
25	28	100.0	979	6	ABU04794 Human exp

26	28	100.0	979	7	ADE57230 Human Pro
27	28	100.0	979	7	ADE56167 Human Pro
28	28	100.0	979	7	ADE57226 Human Pro
29	28	100.0	979	7	ADD45278 Human Pro
30	28	100.0	979	7	ADE57218 Human Pro
31	28	100.0	979	7	ADE57222 Human Pro
32	24	85.7	24	3	AAB30260 CD4+ T-ce
33	19	67.9	19	3	AAB30258 CD4+ T-ce
34	19	67.9	916	2	AAR45779 PTFase PT
35	19	67.9	916	2	AAR45779 PTFase PT
36	19	67.9	961	6	ABU04786 Human exp
37	19	67.9	961	6	ABU04787 Human exp
38	19	67.9	978	8	ADO60054 CRH signa
39	19	67.9	996	5	ABB57041 Mouse isc
40	18	64.3	18	3	AAB30259 CD4+ T-ce
41	17	60.7	17	3	AAB30257 CD4+ T-ce
42	16	57.1	16	3	AAY59322 Tyrosine
43	16	57.1	16	3	AAB30256 CD4+ T-ce
44	13	46.4	13	3	AAB30255 CD4+ T-ce
45	10	35.7	16	3	AAY59337 Tyrosine

ALIGNMENTS

RESULT 1

AAY33360
ID AAY33360 standard; protein; 294 AA.

XX AAY33360;

XX 29-NOV-1999 (first entry)

DE Human islet cell antigen clone ICA-512 protein fragment.

XX Islet cell antigen; ICA; human; pancreas; antibody; IDDM; immunoassay;
KW insulin-dependent type I diabetes mellitus; diagnosis; immunoglobulin;
KW T-cell; B-cell; islet cell destruction; immunoreactivity; disease;
KW immune therapy; stimulating antigen; screening.

XX Homo sapiens.

XX US5955345-A.

PD 21-SEP-1999.

XX 06-JUN-1995; 95US-00468576.

PR 17-FEB-1989; 89US-00312543.

PR 04-DEC-1989; 89US-00441703.

PR 14-JUN-1991; 91US-00715181.

PR 08-JUN-1992; 92US-00672646.

XX 05-MAY-1994; 94US-00239276.

(FARB) BAYER CORP FORMERLY MOLECULAR DIAGNOSTIC.

Rabin DU;

WPI; 1999-560494/47.

N-PSDB; AAZ09922.

Isolated DNA encoding antigens reactive with serum from diabetics useful to develop products for the diagnosis and monitoring of insulin-dependent diabetes mellitus.

Example; Col 85-87; 63pp; English.

XX This invention describes novel isolated and purified DNAs encoding pancreatic islet cell antigens reactive with islet cell antibodies found in sera of patients with insulin-dependent type I diabetes mellitus (IDDM). The polypeptides of the invention are useful as immunoassay reagents in the presymptomatic diagnosis of insulin-dependent type I diabetes mellitus (IDDM). The ability of the polypeptides to bind the

CC antibody binding site on islet cell antibodies (ICAs) also confers
 CC utility in the binding or blocking of human immunoglobulin, T-cells, or B
 CC -cells involved in IDDM. The products can also be used to study the
 CC biological mechanisms involved in islet cell destruction and the
 CC appearance of ICA. The immunoreactivity profile with different antigens
 CC can provide diagnostically significant information concerning the nature
 CC of the disease, e.g. subtypes, the state of the disease, the proximity to
 CC onset of the disease, and the efficacy of therapy, e.g. immune therapy.
 CC The antigens can serve as stimulating antigens for T-cell culture,
 CC permitting significantly improved T-cell cloning, identification, and
 CC growth. The availability of large quantities of pure antigen enables the
 CC development of highly sensitive and specific immunoassays which can be
 CC used to screen the general population of presymptomatic IDDM or a
 CC predisposition to develop IDDM. This sequence represents the human islet
 CC cell antigen from clone ICA-512
 XX
 XX Sequence 294 AA;

Query Match 100.0%; Score 28; DB 2; Length 294;
 Best Local Similarity 100.0%; Pred. No. 1.2e-21;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LAKEWQALCAYQAEPTNCATAGGEGNIK 28
 DB 67 LAKEWQALCAYQAEPTNCATAGGEGNIK 94

RESULT 2

AAAY49854
 ID AAY49854 standard; protein; 294 AA.

XX
 AC AAY49854;

DT 21-JAN-2000 (first entry)

XX Human pancreatic islet cell antigen ICA-512 ATCC 40706 protein.

DE Human; pancreatic islet cell antigen; ICA clone; IDDM; diabetes;
 KW insulin dependent type I diabetes mellitus; diagnosis; detection;
 KW immunoglobulin; T-cell; B-cell; antibody binding.

XX Homo sapiens.

OS US981700-A.

PN 09-NOV-1999.

PD 06-JUN-1995; 95US-00468579.

PF 17-FEB-1989; 89US-00312543.

PR 04-DEC-1989; 89US-00441703.

PR 14-JUN-1991; 91US-00715181.

PR 08-JUN-1992; 92US-00872646.

PR 05-MAY-1994; 94US-00239276.

XX (FARB) BAYER CORP FORMERLY MOLECULAR DIAGNOSTIC.

XX Rabin DU;

XX WPI; 1999-633370/54.

XX N-PSDB; AAZ32341.

XX New pancreatic islet cell antigens, useful in the diagnosis of insulin-
 PT dependent (Type 1) diabetes mellitus.

XX Claim 1; Col 83-86; 61pp; English.

XX The present sequence represents a human pancreatic islet cell antigen
 CC (ICA) that binds with antibodies found in the sera of patients afflicted
 CC with insulin-dependent (Type 1) diabetes mellitus (IDDM). ICAs from the
 CC present invention are encoded by the DNA insert of a recombinant cloning
 CC vehicle selected from ATCC 40550 (AAZ32333, encoding AAY49847), ATCC
 CC 40553 (AAZ32334, encoding AAY49848), ATCC 40554 (AAZ32335, encoding

CC AAY49849), ATCC 40551 (AAZ32336, encoding AAY49850), ATCC 40552
 CC (AAZ32337, encoding AAY49851), ATCC 40703 (AAZ32338, encoding AAY49852),
 CC ATCC 40704 (AAZ32339, encoding AAY49853), ATCC 40705 (AAZ32340), ATCC
 CC 40706 (AAZ32341, encoding AAY49854) and ATCC 75030 (AAZ32342, encoding
 CC AAY49855). ICA proteins and their peptide fragments can be used in the
 CC diagnosis of IDDM and in detecting or blocking human immunoglobulin, T-
 CC cells or B-cells involved in IDDM

XX Sequence 294 AA;

Query Match 100.0%; Score 28; DB 2; Length 294;

Best Local Similarity 100.0%; Pred. No. 1.2e-21;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LAKEWQALCAYQAEPTNCATAGGEGNIK 28

DB 67 LAKEWQALCAYQAEPTNCATAGGEGNIK 94

RESULT 3

AAW80486
 ID AAW80486 standard; protein; 294 AA.

XX
 AC AAW80486;

DT 01-FEB-1999 (first entry)

DE Islet cell antibody antigen encoded by clone ICA-512.

XX Pancreatic islet cell antibody antigen; ICA antigen; diagnosis;

KW treatment; insulin dependent diabetes mellitus; IDDM; autoantibody;

KW autoimmune disease; human.

XX Homo sapiens.

XX US5840836-A.

XX 24-NOV-1998.

XX 05-MAY-1994; 94US-00239276.

XX 17-FEB-1989; 89US-00312543.

XX 04-DEC-1989; 89US-00441703.

XX 14-JUN-1991; 91US-00715181.

XX 08-JUN-1992; 92US-00872646.

XX (FARB) BAYER CORP.

XX Rabin DU;

XX WPI; 1999-034118/03.

XX N-PSDB; AAV63560.

XX Polypeptide antigens from pancreatic islet cells - useful for the
 PT detection and treatment of insulin dependent diabetes mellitus.

XX Claim 1; Col 51-54; 47pp; English.

XX The present sequence represents a pancreatic islet cell antibody (ICA)
 CC antigen. The proteins are used for the diagnosis and treatment of insulin
 CC dependent (Type 1) diabetes mellitus (IDDM). They are recognised by
 CC autoantibodies of the disease as IDDM is now thought to be an autoimmune
 CC disease. They were identified from probing of IDDM serum with human cDNA
 CC libraries expressing protein fragments

XX Sequence 294 AA;

Query Match 100.0%; Score 28; DB 2; Length 294;

Best Local Similarity 100.0%; Pred. No. 1.2e-21;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LAKEWQALCAYQAEPTNCATAGGEGNIK 28